

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

UM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 15:55:18 ; Search time 49.1 seconds

(without alignments)
4030.649 Million cell updates/sec

Title: US-08-799-910-9
Perfect score: 1228
Sequence: 1 ATGTCTCAGCTCGCAGCTG.....AAAAAAAAAACTGAG 1228

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6.COMB.seq:*
4: /cgn2_6/prodata/2/ina/PCTUS.COMB.seq:*
5: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1228	100.0	1228	3	US-08-826-246-9
2	1228	100.0	1228	3	US-08-944-495-9
3	1228	100.0	1228	3	US-09-126-640-5
4	49.8	4.1	7218	1	US-08-232-463-14
5	43.8	3.6	7218	1	US-08-232-463-14
6	42	3.4	8438	1	US-07-945-283-1
7	41.2	3.4	530	2	US-08-758-662-4
8	41.2	3.4	4524	2	US-08-845-998-7
9	41.2	3.4	4524	3	US-09-206-537-7
10	40.6	3.3	15231	3	US-09-128-155-16
11	40	3.3	2847	1	US-08-087-007-2
12	40	3.3	2847	3	US-08-483-433-2
13	40	3.3	2847	4	PCT-US92-05920-2
14	39.8	3.2	1120	3	US-09-030-613-1
15	38.6	3.1	6354	3	US-09-058-389A-5
16	38.4	3.1	6803	3	US-08-665-259-19
17	38.4	3.1	6803	3	US-08-762-500-19
18	37.6	3.1	170	3	US-09-058-389A-13
19	37.6	3.1	4425	2	US-08-749-169A-1
20	37.6	3.1	4425	2	US-09-130-032A-1
21	37.2	3.0	1420	2	US-08-909-965C-3
22	37.2	3.0	371	1	US-08-664-596B-25
23	37.2	3.0	371	2	US-08-739-775-3
24	37.2	3.0	20235	1	US-07-642-734C-3
25	37.2	3.0	20235	3	US-08-439-009A-3
26	36.6	3.0	1931	2	US-09-130-114-2
27	36.4	3.0	702	1	US-08-458-568A-3
28	36.4	3.0	1841	2	US-08-820-521-1

29	36.4	3.0	3181	3	US-09-135-021-1	Sequence 1, Appl 1
C 30	36.4	3.0	3182	3	US-09-135-021-5	Sequence 5, Appl 1
C 31	36.4	3.0	12001	1	US-08-458-568A-11	Sequence 11, Appl 1
32	36.2	2.9	1833	2	US-08-403-892D-6	Sequence 6, Appl 1
33	36.2	2.9	1833	2	US-08-510-646B-6	Sequence 6, Appl 1
34	36	2.9	18318	1	US-08-414-926A-6	Sequence 6, Appl 1
35	36	2.9	18318	2	US-08-926-922-6	Sequence 6, Appl 1
36	36	2.9	18318	3	US-09-253-682-6	Sequence 6, Appl 1
37	35.8	2.9	2728	3	US-09-188-930-213	Sequence 213, App
38	35.8	2.9	12001	1	US-08-458-568A-11	Sequence 11, Appl
C 39	35.6	2.9	3978	3	US-08-726-214-1	Sequence 1, Appl 1
40	35.4	2.9	1680	1	US-08-140-729A-2	Sequence 2, Appl 1
41	35.4	2.9	1680	1	US-08-546-666-2	Sequence 2, Appl 1
42	35.4	2.9	1680	2	US-08-916-745-2	Sequence 2, Appl 1
43	35.4	2.9	1680	2	US-09-042-929-2	Sequence 2, Appl 1
44	35.4	2.9	1680	2	US-08-546-661-2	Sequence 2, Appl 1
45	35.4	2.9	1680	2	US-09-042-960-2	Sequence 2, Appl 1

ALIGNMENTS

RESULT 1
US-08-826-246-9
Sequence 9, Application US/08826246
Patent No. 6048709
GENERAL INFORMATION:
APPLICANT: Faib, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826, 246
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799, 910
FILING DATE: 13-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011, 787
FILING DATE: 16-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-078-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1228 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...468
OTHER INFORMATION:

US-08-826-246-9

Query Match	100.0%	Score 1228;	DB 3;	Length 1228;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1228;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	1	ATATGTCACATCTGGACAGCTGCCACCCGAGCATGAGACATCTGCAAGGCCACCCGCGGC	60
Pb	1	ATGTGTACTCTTGGACAGTGTCCACCCGACATGACCATCTGTGTCAAGGCCCGACCCGCGGC	60
OY	61	CCCTTCACCATTCGCCGGGACCCCGCGGGCTTCGGTCTCCAGATTCCTCACCTTCGACCT	120
Pb	61	CCCTTCACCATTCGCCGGGACCCCGCGGGCTTCGGTCTCCAGATTCCTCACCTTCGACCT	120
OY	121	CTCCCGGAGCCCGCACGGGCGCTGTGCGGGGGCCCGCACGGCTCTGTGGCGGACCCGAAG	180
Pb	121	CTCCCGGAGCCCGCACGGGCGCTGTGCGGGGGCCCGCACGGCTCTGTGGCGGACCCGAAG	180
OY	181	CGCAGCCGCGAGGGTCTCTACCTTCGATGTGTCCGGCGCCACAGCTCCAGTGTAGAGAACCG	240
Pb	181	CGCAGCCGCGAGGGTCTCTACCTTCGATGTGTCCGGCGCCACAGCTCCAGTGTAGAGAACCG	240
OY	241	AACCCAGCCAAAGGCTTCTTTCGTGTCTCACCATCTGTCTTGTGCCAATCTCTATG	300
Pb	241	AACCCAGCCAAAGGCTTCTTTCGTGTCTCACCATCTGTCTTGTGCCAATCTCTATG	300
OY	301	GCTCAGAGAGGTGTGCCGGGCGCCCTGCTCCAGAGAGAGCCCTTAACGCGGCATCTCGT	360
Pb	301	GCTCAGAGAGGTGTGCCGGGCGCCCTGCTCCAGAGAGAGCCCCCTTAACGCCCATCTCTG	360
OY	361	CGCGCCACCCCTGTGTCCCGCGTGTCTCGAGGCCCTTAATGTGACTTCGAGACCCCTCGAC	420
Pb	361	CGCGCCACCCCTGTGTCCCGCGTGTCTCGAGGCCCTTAATGTGACTTCGAGACCCCTCGAC	420
OY	421	TACCGTGTGAGACCGACACACTTTCGTCCAGCAAGCCCGCGCGCTTGTAACTGTGACATC	480
Pb	421	TACCGTGTGAGACCGACACACTTTCGTCCAGCAAGCCCGCGCTTGTAACTGTGACATC	480
OY	481	CTCGACTCTCCCAAAAAGATTCGAAACCAACAAAGAAACCAACAGCGCTACTGTGTGCG	540
Pb	481	CTCGACTCTCCCAAAAAGATTCGAAACCAACAAAGAAACCAACAGCGCTACTGTGTGCG	540
OY	541	CGAAGCGCTATCCCACTGTGGACATTCGAGGCACTTGTAACTCAAGACACTACAGCGGA	600
Pb	541	CGAAGCGCTATCCCACTGTGGACATTCGAGGCACTTGTAACTCAAGACACTACAGCGGA	600
OY	601	GAGCCACCCCGTGTGTGAGCGGAGACGAGCGCACAGACACCGACGCGCATAGACACC	660
Pb	601	GAGCCACCCCGTGTGTGAGCGGAGACGAGCGCACAGACACCGACGCGCATAGACACC	660
OY	661	GAGTCACAGCCACGTGGGCTAGCGCCGGTGGGAAGAGAGCTGTTAATTAATTTCT	720
Pb	661	GAGTCACAGCCACGTGGGCTAGCGCCGGTGGGAAGAGAGCTGTTAATTAATTTCT	720
OY	721	TATTGCTCTCAATTAATATTATATGATATTATATGATCGATCGTAGGATGATGAGATGT	780
Pb	721	TATTGCTCTCAATTAATATTATATGATATTATATGATCGATCGTAGGATGATGAGATGT	780
OY	781	GTACCTTAATTAATTAATTTAACTTATGCAAGGGTGTGACATCTTCCCTGTCTGTTAAATCC	840
Pb	781	GTACCTTAATTAATTAATTTAACTTATGCAAGGGTGTGACATCTTCCCTGTCTGTTAAATCC	840
OY	841	AGGCTCTCTGTTATTATTATGAGCTTTGTGTGAGATGTGTGTAACAGAGACACTGGAACTGC	900
Pb	841	AGGCTCTCTGTTATTATTATGAGCTTTGTGTGAGATGTGTGTAACAGAGACACTGGAACTGC	900
OY	901	GCCAAATACAGAGAGAAATGTGGAGAGACTCGCGTGGGGAGAGAGCTCCGCTGTGATG	960
Pb	901	GCCAAATACAGAGAGAAATGTGGAGAGACTCGCGTGGGGAGAGAGCTCCGCGTGTGATG	960
OY	961	AAGTCTGCTGCTGGTCTGAATTTAGAGAGGTGACTGACTCTCCAGACTCTCAACTCCG	1020
Pb	961	AAGTCTGCTGCTGGTCTGAATTTAGAGAGGTGACTGACTCTCCAGACTCTCAACTCCG	1020

```

Db 961 AAGTCGTGCTGGCCCTCTAGATGTTAGAGAGTGACATCCATCCACCACTATCACAACTCGC 1020
QY 1021 TCTGCTCTCTGTGAGACCTTGCGGGAGACCAFTTGGATGAGATCCGTGAGATCTCTTCA 1080
Db 1021 TCTGTCTACTGTGTGAGACTTGGGGAGACCATTAGAGATGAGATCCGTGAGATCTCTTCA 1080
QY 1081 TCTTCTTGAAGTCGCGCTTTAGAGCTGGCTGCGAGGTACAGCGTTGGCGCTGCTCCGCTCT 1140
Db 1081 TCTTCTTGAAGTCGCGCTTTAGAGGTGGCTGCGAGGTAGAGGGTTGGGTGGGCTGT 1140
QY 1141 CACGAGGCGACTGTGGAATGGCCTAGATGTCTCTGTACACACAATTAATGTATTAC 1200
Db 1141 CACGAGGCGACTGTGGAATGCCCTAGATGTCTCTGTGAACACAATTAATGTATTAC 1200
QY 1201 TGTCAAAAAAAAAAAAAAAAAAACTGAG 1228
Db 1201 TGTCAAAAAAAAAAAAAAAAAAACTGAG 1228

RESULT: 2
US-08-944-495-9
Sequence 9, Application US/08944495
Patent No. 6087477
GENERAL INFORMATION:
APPLICANT: Fald, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
TITLE OF INVENTION: CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,495
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-067-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8659741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1228 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..468
OTHER INFORMATION:
US-08-944-495-9

Query Match 100.0% Score 1228: DB 3: Length 1228:
Best Local Similarity 100.0%: Pred NO. 0:
Matches 1228: Conservative 0: Mismatches 0: Indels 0: Gaps 0

```



```

QY 1 ATGTGACTCTGTCAGACCTGCAACCCGACCATGACATCTCTGGAGGCCCCGACCCGCGCC 60
DB 1 ATGTGACTCTGTCAGACCTGCAACCCGACCATGACATCTCTGGAGGCCCCGACCCGCGCC 60
QY 61 CCCCTCACCATCTCCCGGACCCCGGCGGGCTCCGGTCCCTGAGATCTTTCACCTTTCAGACCT 120
DB 61 CCCCTCACCATCTCCCGGACCCCGGCGGGCTCCGGTCCCTGAGATCTTTCACCTTTCAGACCT 120
QY 121 CTCGCGGAGCGCCGACCGCCGCTGCGCGGCGCCGACCGCCCTCTCGCGCGGACCGAAG 180
DB 121 CTCGCGGAGCGCCGACCGCGCCGCTGCGCGGCGCCGACCGCCCTCTCGCGCGGACCGAAG 180
QY 181 CGCAGCGCAGAGGCTTCTACCTGAGTGTCCGGCGCGAGTCCGAGTCCGAGGAAACCG 240
DB 181 CGCAGCGCAGAGGCTTCTACCTGAGTGTCCGGCGCGAGTCCGAGTCCGAGGAAACCG 240
QY 241 AACCCAGCCAAAGGCTTCTCTTCTGCTGCTCAGCATCTCTTCTCCGAGATCTGATG 300
DB 241 AACCCAGCCAAAGGCTTCTCTTCTGCTGCTCAGCATCTCTTCTCCGAGATCTGATG 300
QY 301 GCTGAGACGCGTGTCCCGCGCGCCGCGCCCTCCAGAGACCGCCCTTAACGCGCATCCCTG 360
DB 301 GCTGAGAGGCGTGTCCCGCGCGCCGCGCCCTCCAGAGAGCGCCCTTAACGCGCATCCCTG 360
QY 361 GCGCCACCCCTGTGTCTCCCGCTCTGAGCCCTTAACTGACTTCCGAGGCGCTCGGAC 420
DB 361 GCGCCACCCCTGTGTCTCCCGCTCTGAGCCCTTAACTGACTTCCGAGGCGCTCGGAC 420
QY 421 TACGCTGTGACCTGACACTTCTCTCAGCAACACCGCGCGCCCTTCTAATCTGACTC 480
DB 421 TACGCTGTGACCTGACACTTCTCTCAGCAACACCGCGCGCCCTTCTAATCTGACTC 480
QY 481 CCCGCACTCCCAAAAGATCCGAAACCAACCAAGAAACACACGAGCTACTGCTGG 540
DB 481 CCCGCACTCCCAAAAGATCCGAAACCAACCAAGAAACACACGAGCTACTGCTGG 540
QY 541 CGAGAGCGTATCCCACTGAGGACTTCCGAGGCACTTGAACCTCAGAACACTACAGCGA 600
DB 541 CGAGAGCGTATCCCACTGAGGACTTCCGAGGCACTTGAACCTCAGAACACTACAGCGA 600
QY 601 GACGCGACCCGCTGCTTGAAGCGGAGACGAGCGCACAGAGACCGAGCGCATATGAGACC 660
DB 601 GACGCGACCCGCTGCTTGAAGCGGAGACGAGCGCACAGAGACCGAGCGCATATGAGACC 660
QY 661 GAGGCGACCCGAGTGGGCTAGGCGCGTGGGAAAGGAGACGCTCTAATTTATTTCT 720
DB 661 GAGGCGACCCGAGTGGGCTAGGCGCGTGGGAAAGGAGACGCTCTAATTTATTTCT 720
QY 721 TATTCCTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTCT 780
DB 721 TATTCCTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTCT 780
QY 781 GACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 840
DB 781 GACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 840
QY 841 AGCTCTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 900
DB 841 AGCTCTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 900
QY 901 GGCAGAGTGAAGAAATGGGAGGACTCGGCTGGGAGGAGGAGCTTCGCGCTGGATG 960
DB 901 GGCAGAGTGAAGAAATGGGAGGACTCGGCTGGGAGGAGGAGCTTCGCGCTGGATG 960
QY 961 AAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020
DB 961 AAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020
QY 1021 TCTGCTCTGCTGAGACTTCGCGGAGCATTAAGAAATGAGATCCGAGATCTCTTCA 1080
DB 1021 TCTGCTCTGCTGAGACTTCGCGGAGCATTAAGAAATGAGATCCGAGATCTCTTCA 1080

```

```

QY 1081 TCTTCTTGAGTGCCTTTAAGGTGGCTGCGAGGTAGAGGTTGGGGTGTGTGACCTGT 1140
DB 1081 TCTTCTTGAGTGCCTTTAAGGTGGCTGCGAGGTAGAGGTTGGGGTGTGTGACCTGT 1140
QY 1141 CAGGAGCGAGCTGCGAGATCGCCTAGTATGCTGTGAGACACAAATAAATGATTTAC 1200
DB 1141 CAGGAGCGAGCTGCGAGATCGCCTAGTATGCTGTGAGACACAAATAAATGATTTAC 1200
QY 1201 TGTCAAAAAAAAAAAAAAAAAAATCTGAC 1228
DB 1201 TGTCAAAAAAAAAAAAAAAAAAATCTGAC 1228

```

RESULT 3

```

US-09-126-640-5
; Sequence 5, Application US/09126640A
; Patent No. 6099823
; GENERAL INFORMATION:
; APPLICANT: FALD, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/126,640A
; EARLIER APPLICATION NUMBER: 08/870,434
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 08/799,910
; EARLIER FILING DATE: 1997-02-13
; EARLIER APPLICATION NUMBER: 60/011,787
; EARLIER FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FASTED for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1228
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-126-640-5

```

```

Query Match 100.0%; Score 1228; DB 3; Length 1228;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGTGACTCTGTCAGACCTGCAACCCGACCATGACATCTCTGGAGGCCCCGACCCGCGCC 60
DB 1 ATGTGACTCTGTCAGACCTGCAACCCGACCATGACATCTCTGGAGGCCCCGACCCGCGCC 60
QY 61 CCCCTCACCATCTCCCGGACCCCGGCGGGCTCCGGTCCCTGAGATCTTTCACCTTTCAGACCT 120
DB 61 CCCCTCACCATCTCCCGGACCCCGGCGGGCTCCGGTCCCTGAGATCTTTCACCTTTCAGACCT 120
QY 121 CTCGCGGAGCGCCGACCGCCGCTGCGCGGCGCCGACCGCCCTCTCGCGCGGACCGAAG 180
DB 121 CTCGCGGAGCGCCGACCGCCGCTGCGCGGCGCCGACCGCCCTCTCGCGCGGACCGAAG 180
QY 181 CGCAGCGCAGAGGCTTCTACCTGAGTGTCCGGCGCGAGTCCGAGTCCGAGGAAACCG 240
DB 181 CGCAGCGCAGAGGCTTCTACCTGAGTGTCCGGCGCGAGTCCGAGTCCGAGGAAACCG 240
QY 241 AACCCAGCCAAAGGCTTCTCTTCTGCTGCTCAGCATCTCTTCTCCGAGATCTGATG 300
DB 241 AACCCAGCCAAAGGCTTCTCTTCTGCTGCTCAGCATCTCTTCTCCGAGATCTGATG 300
QY 301 GCTGAGAGCGTGTCCCGCGCGCCCTCCAGAGAGCGCCCTTAACGCGCATCCCTG 360
DB 301 GCTGAGAGCGTGTCCCGCGCGCCCTCCAGAGAGCGCCCTTAACGCGCATCCCTG 360
QY 361 GCGCCACCCCTGTGTCTCCCGCTCTGAGCCCTTAACTGACTTCCGAGGCGCTCGGAC 420
DB 361 GCGCCACCCCTGTGTCTCCCGCTCTGAGCCCTTAACTGACTTCCGAGGCGCTCGGAC 420
QY 421 TACGCTGTGACCTGACACTTCTCTCAGCAACACCGCGCGCCCTTCTAATCTGACTC 480
DB 421 TACGCTGTGACCTGACACTTCTCTCAGCAACACCGCGCGCCCTTCTAATCTGACTC 480

```

421 taagctcgaactcagcacttccctccagcaacacccgcgccttcaatgtgactc 480
QY 401 CCCGCACTCCCAAAAGATCCGAAAACACAAAGAACACGAGCGGTACCTGGTCG 540
Db 481 cccgcactcccaaaaagaaacgcgaaaacacaaagaaacacacgagcgtacccg 540
QY 541 CGAGAGCGCTATCCCACTGGGACTTCCGAGCGCAACTTGACACTGACACTAAGCCGA 600
Db 541 cgaagcgtatcccaactcgaactcgaagcaactgaactcgaactcgaactcga 600
QY 601 GAGCGCACCCGCTCTTACGCGGAGCGGACGACGACGACGACGACGACGACGACG 660
Db 601 gacgcacccggtctcgaagcgcgagccgagccgagccgagccgagccgagccg 660
QY 661 GAGCAGACCCGAGCTGGGCTAGGCGCGGTGGAGAGAGAGAGAGAGAGAGAGAG 720
Db 661 gagagacacccagcctgggagccgagccgagccgagccgagccgagccgagcc 720
QY 721 TATTCCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
Db 721 taatgctcctaaatlaattatlaattatlaattatlaattatlaattatlaatt 780
QY 781 GTACGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
Db 781 gtaagcaatcattcattcattcattcattcattcattcattcattcattcattc 840
QY 841 AGGTCTTGTGATTTATTTAGACTTTGTGAGACTGTTGAGAGAGAGAGAGAGAG 900
Db 841 aggtctcctgatttattgactttgtgagactgttgagagagagagagagagag 900
QY 901 GCGCAAGTACGAGAAAGATCGGAGAGACTCGGCTGGGAGAGAGAGAGAGAGAG 960
Db 901 ggcgaagtagagagaaatgggagagactcgggagagagagagagagagagag 960
QY 961 AGGTGAGTGGGCTGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1020
Db 961 aagtgagtggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1020
QY 1021 TCTGCTACTGTGTAGACTTGGCGGACCATTAAGAAATGAGATCGGTGATCTTCCA 1080
Db 1021 tctgctactgtgtgagacttggcggaacattaggaatcgaagatcgaagatcga 1080
QY 1081 TCTTCTTACAGTCGCTTTAGAGCTGGCTCGGACGTAGAGAGCTTGGCGCTCT 1140
Db 1081 tcttcttgaagtcgacttgaagtcgacttgaagtcgacttgaagtcgacttga 1140
QY 1141 CACGAGCAGCTGTGAGATCGCTAGTATGTTCTGTGACACAAATAAATTGATTTAC 1200
Db 1141 cagcgaagcagctgtcgaagtcgacttgaagtcgacttgaagtcgacttgaag 1200
QY 1201 TGTCAAAAAGAAAAAAGAACTCGAG 1228
Db 1201 tgtcaaaaaaagaaaaaagaaactcga 1228

RESULT 4
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHREIFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)883-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpL-F15
US-08-232-463-14

Query Match 4.18; Score 49.8; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 0.0012;
Matches 6; Conservative 220; Mismatches 147; Indels 0; Gaps 0;

QY 97 CCGAGATCTTACCTTGCAGCCCTCTCCGAGAGCCCGAGCGCCCTCCCGGCCCC 156
Db 1059 CTGGGATTT 1118
QY 157 AGCGCTCTCGCGGACCGAAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 216
Db 1119 TT 1178
QY 217 GCGCAGCTCGAGTGCAGAACCGAACCGAACCGAACCGAACCGAACCGAACCG 276
Db 1179 TT 1238
QY 277 ATCGTCTTCCGATCTGATGCTGTAGAGAGCTGCGCGCGCCCTCCCTCCAGAG 336
Db 1239 TT 1298
QY 337 GAGCGCCCTAAGCGCGATCCCTGGCGCCACCCCTGTGCTCCCTGAGCCCTTT 396
Db 1299 TT 1358
QY 397 AATCTGACTCGAGCCCTCGAGACTAGCTGTGAGACTGAGACTTCTCCAGCAACAC 456
Db 1359 TT 1418
QY 457 CCGCGCGCTCTCT 469
Db 1419 TTTTTTTTTTTTTT 1431

RESULT 5
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHREIFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:

[illegible][illegible]

```

RESULT      9
US-09-206-537-7/c
: Sequence 7, Application US/09206537
: Patent No. 6130052
: GENERAL INFORMATION:
: APPLICANT: Van Baren, Nicolas
: APPLICANT: Coulle, Pierre G.
: APPLICANT: De Smet, Charles
: APPLICANT: Lucas, Sophie
: APPLICANT: Boon, Thierry
: TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
: STREET: 600 Atlantic Avenue
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/206,537
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/845,998
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Van Amsterdam, John R.
: REGISTRATION NUMBER: 40,212
: REFERENCE/DOCKET NUMBER: L0461/7008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)720-3500
: TELEFAX: (617)720-2441
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4524 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 174..1433
:
US-09-206-537-7

Query Match      3.4%; Score 41.2; DB 3; Length 4524;
Best Local Similarity 53.8%; Pred. No. 0.18;
Matches 85; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY      35  CCAATCTTCAGAGCCCGGACCCCGGCCCCCTCCACCATCCGGGAGCCCGGCGGCTCCG 94
           11 11 11111 111 111 111 111 111 111 111 111 111 111 111 111
Db      697  CCGGCCCCCGGCGCGCCGCGCCGCTCCCGGCGCGCGCGCGCGCGCGCGCGCGCG 638

OY      95  GTTCGAGATTTCACCTTGACCTTCGCCGAGCCCGGCAAGCGCGCCCTCGCGGGCGG 154
           11 11 11111 111 111 111 111 111 111 111 111 111 111 111 111
Db      637  CCGCGCGGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 578

OY      155 CAGAGCGCTTCAGCGGCGGCGACCGAAAGCGCAGCGCAGG 192
           11 11111 1111 11 11111 11 11
Db      577  CCGCGCGCGCGCGCGCGCGCGCGCATGAGCGCGAGCGAGCG 540

RESULT      10
US-09-128-155-16

```

```

Query Match      3.3% Score 40; DB 1; Length 2847;
Best Local Similarly 51.1%; Pred No. 0.3;
Matches 94; Conservative 0; Mismatches 90; Indels 0; Gaps 0

OY       326 TGCCTCCAGACGACGCCCTTAACGGCCGATCCTTGCGGCCCAACCCCTGTGCCCCCTCC 385
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db        505 TGAGTCAGAGAGGGTGTCCAGAGACTCTCTCTCTCCCTCCCTCCCACTCTCCCAGAGTC 564

OY       386 TCAGACCCTTTAATCTGTACTCTGGAGCCCTCGACTAAGCTTTGGAACTTCACACTTTCC 445
          . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db        565 TAGCCCCCCGGGATGACACCCGGAGCCCTTGACGCGACCTCTGACCAACAACAACCC 624

OY       446 TCCAGCAACACCGCGCCCTCTTAACTGTGACTCCCGACTCCCAAAAAGAAATCCGA 505
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db        625 TACTCCACCCGCTCTTGTATTGTCCACCCCTGTGTGAGAGCAAGACCCCAAGCCACAGCCCGC 684

OY       506 AAAA 509
          |||
Db        685 CCAA 688

RESULT 12
US-08-4B3-433-2
Sequence 2, Application US/08483433
Patent No. 6100443
GENERAL INFORMATION:
```

[illegible]

OY 506 AAAA 509
DB 685 CCAA 688

RESULT 13

PCT-US92-05920-2

Sequence 2, Application PC/TUS9205920

GENERAL INFORMATION:

APPLICANT: Sims, Peter J.

APPLICANT: Botchwell, Alfred L.M.

APPLICANT: Elliott, Eileen A.

APPLICANT: Flavell, Richard A.

APPLICANT: Madri, Joseph

APPLICANT: Rollins, Scott

APPLICANT: Bell, Leonard

APPLICANT: Squinto, Stephen

TITLE OF INVENTION: Universal Donor Cells

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kilpatrick & Cody

STREET: 1100 Peachtree Street, Suite 2800

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.

ZIP: 30309-4530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/05920

FILING DATE: 19920714

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: OMRF135

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-815-6500

TELEFAX: 404-815-6555

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2847 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

LIBRARY: Genbank HUMDAF: HUMDAF1

CLONE: Human DAF cDNA

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..819

OTHER INFORMATION: /note="HUMDAF1 (Promotor and 5'

OTHER INFORMATION: end of Exon 1, genomic sequence)"

PCT-US92-05920-2

Query Match 3.38; Score 40; DB 4; Length 2847;

Best Local Similarity 51.18; Pred. No. 0.3;

Matches 94; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

OY 326 TGCCTCCAGACGACCCCTTAACGCGCATCCCTGGCCGCCACCCCTGTGTCGCCGTCG 385
DB 505 TGACTCCAGAGCGCTGTCCAGAGACCTCTCTCCCTCCCTCCACCTCCGAGCAGC 564
OY 386 TCGAGCCCTTAATCTGACTGAGAGCCCTCGACTACGCTCTGAGACCTGAGCACTTCC 445

DB 565 TAGGGCCCCGGGGTATGAGAGCCGCTGTGAGCCGACCTCTACCAACCAACACCC 624

OY 446 TCCAGACACACCGCGCCCTTCTACTGTGACTCCCGACACTCCCAAAAGAAATCCGA 505
DB 625 TACTCCACCCGCTGTGTTGTCACACCTGTGTGAGCAGAGACCCAGCCAGACCCGCG 684

OY 506 AAAA 509
DB 685 CCAA 688

RESULT 14

US-09-030-613-1/C

Sequence 1, Application US/09030613

Patent No. 6083706

GENERAL INFORMATION:

APPLICANT: Florjanczyk, Robert Z.

APPLICANT: Baird, J. Andrew

TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/030.613

FILING DATE: 25-FEB-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: No. 6083706tenburg Ph.D., Carol

REGISTRATION NUMBER: 39,317

REFERENCE/DOCKET NUMBER: 760100.418C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 682-6031

TELEFAX: (206) 682-4900

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1120 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-030-613-1

Query Match 3.24; Score 39.8; DB 3; Length 1120;

Best Local Similarity 51.48; Pred. No. 0.22; Mismatches 87; Indels 0; Gaps 0;

Matches 92; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

OY 39 CCTGACGCCCCGAGCCCGCCCTTCACCAATCCGGGAGCCCGGCGGCTCCGTCG 98
DB 362 CCTGCGGGGCGCGCGCGGATCCCGAGCTGTGAGCGCGCGGAGCGCGCGCTGCC 303
OY 99 TGAGATCTTACCTTCAGACCTCTCCGAGCGCCGACGAGCCCTGTGCGGCGGCCAG 158
DB 302 CCGGCCCGCGCGCGCGCTTCGACCGCTCCGCGGCGACGCGCCCGCCCGCCCGCAG 243
OY 159 CGCCTCTGCGGCGACCGAAAGCGAGCGAGGTTCTTACCTTCGAGTGTGCGGC 217
DB 242 CCTTCGCCCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 184

RESULT 15

US-09-058-389A-5

Sequence 5, Application US/09058389A

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using SW model

Run on: January 30, 2001, 16:08:34 ; Search time 77.57 Seconds

5947.064 Million cell updates/sec

Title: US-08-799-910-9

Perfect score:

Sequence: 1 ATGTCCTCACTCTCGCAGCTG.....AAAAAACTCGAC 1228

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 480022 segs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1	/cgn2_2/gcgdata/gemseq/gemseqn/NA1980.DAT *
2	/cgn2_2/gcgdata/gemseq/gemseqn/NA1981.DAT *
3	/cgn2_2/gcgdata/gemseq/gemseqn/NA1982.DAT *
4	/cgn2_2/gcgdata/gemseq/gemseqn/NA1983.DAT *
5	/cgn2_2/gcgdata/gemseq/gemseqn/NA1984.DAT *
6	/cgn2_2/gcgdata/gemseq/gemseqn/NA1985.DAT *
7	/cgn2_2/gcgdata/gemseq/gemseqn/NA1986.DAT *
8	/cgn2_2/gcgdata/gemseq/gemseqn/NA1987.DAT *
9	/cgn2_2/gcgdata/gemseq/gemseqn/NA1988.DAT *
10	/cgn2_2/gcgdata/gemseq/gemseqn/NA1989.DAT *
11	/cgn2_2/gcgdata/gemseq/gemseqn/NA1990.DAT *
12	/cgn2_2/gcgdata/gemseq/gemseqn/NA1991.DAT *
13	/cgn2_2/gcgdata/gemseq/gemseqn/NA1992.DAT *
14	/cgn2_2/gcgdata/gemseq/gemseqn/NA1993.DAT *
15	/cgn2_2/gcgdata/gemseq/gemseqn/NA1994.DAT *
16	/cgn2_2/gcgdata/gemseq/gemseqn/NA1995.DAT *
17	/cgn2_2/gcgdata/gemseq/gemseqn/NA1996.DAT *
18	/cgn2_2/gcgdata/gemseq/gemseqn/NA1997.DAT *
19	/cgn2_2/gcgdata/gemseq/gemseqn/NA1998.DAT *
20	/cgn2_2/gcgdata/gemseq/gemseqn/NA1999.DAT *
21	/cgn2_2/gcgdata/gemseq/gemseqn/NA2000.DAT *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
	1	1228	100.0	1228	21	250711	Nucleotide sequence
	2	1226.4	99.9	1228	21	7904471	human Fc6b605 gene
c	3	420.8	39.3	427	21	260483	human colon cancer
	4	371.6	30.3	411	20	986227	EST clone 0417. H
c	5	167.2	13.6	193	20	989271	EST clone Hy66. H
	6	151	12.3	297	15	077554	Human genome fragm
	7	70.4	5.7	267	15	077534	Human genome fragm
c	8	57.4	4.7	1000	21	A02484	Human colon cancer
c	9	55.2	4.5	1218	21	A02488	Human colon cancer
c	10	52.2	4.3	1337	20	217263	Human gene express
c	11	49.6	4.0	1126	21	A02538	Human colon cancer
c	12	48.2	3.9	53	16	T20789	Human gene signatu

C 13	47.8	3.9	42.28	21	Z51683	Human cyclic nucle
C 14	47.2	3.6	114955	20	X53491	Human adenosine A1
C 15	44.8	3.6	1593	21	A02504	Human colon cancer
C 16	44.2	3.6	2275	21	A34740	Human adenosine re
C 17	44.2	3.6	4475	21	A34741	Human adenosine re
C 18	42.8	3.5	1127	21	A02477	Human colon cancer
C 19	42.4	3.5	4356	14	C03753	Cardiac adenylyl C
C 20	42.4	3.5	4356	16	C09540	Cardiac adenylyl C
C 21	42	3.4	8438	15	C073500	DNA encoding Pseud
C 22	42	3.4	34094	20	Z30163	Complete nucleoid
C 23	41.8	3.4	1459	21	A02558	Human colon cancer
C 24	41.6	3.4	989	21	A02559	Human colon cancer
C 25	41.4	3.4	1275	11	C01630	Chicken beta-actin
C 26	41.4	3.4	1344	12	C01318	Expression vector
C 27	41.4	3.4	1345	11	C03062	Hybrid promoter of
C 28	41.2	3.4	1102	20	Z16993	Human gene express
C 29	41.2	3.4	4534	20	V33912	Nucleotide sequenc
C 30	41	3.3	114955	20	X53491	Human adenosine A1
C 31	40.8	3.3	1966	20	X61220	Mouse DNA demethyl
C 32	40.6	3.3	1017	20	Z17058	Human gene express
C 33	40.4	3.3	3300	20	Z14922	Human gene express
C 34	40.4	3.3	2179	15	C07317	Alpha 2, 3-sialyl
C 35	40.4	3.3	2232	15	C077739	Human alpha-2,3-si
C 36	40	3.3	1523	20	Z174507	Human gene express
C 37	40	3.3	1523	20	Z174507	Human gene express
C 38	39.8	3.2	1110	19	V60339	CDNA sequence of f
C 39	39.8	3.2	2188	20	Z027504	Human ovarian tumo
C 40	39.8	3.2	3198	20	X02974	Human IL-1ra BAC C
C 41	39.6	3.2	954	20	X80342	DNA encoding a hum
C 42	39.6	3.2	2661	20	X58439	Thermophilus therm
C 43	39.2	3.2	1804	20	X61218	Human DNA demethyl
C 44	38.8	3.2	749	20	Z16153	Human gene express
C 45	38.8	3.2	1312	20	Z17264	Human gene express

ALIGNMENTS

RESULT	1
ID	250711
ID	250711 standard; DNA; 1228 BP.
XX	
AC	250711;
XX	
DT	31-MAY-2000 (first entry)
XX	
DE	Nucleotide sequence of human fcdh605 gene.
XX	
KW	fcdh605 gene; human; cardiovascular disease; oncogenic disorder;
KW	diabetic retinopathy; fibroproliferative disorder; arteriosclerosis;
KW	TGF-beta signalling pathway; TGF; Transforming growth factor;
KW	pancreatic cancer; angiogenesis; inflammation; tumour growth;
KW	vascularisation; cytosolic; antidiabetic; ophthalmological; ds.
XX	
OS	Homo sapiens.
XX	
EI	Key
EI	Location/Qualifiers
CDS	.1..471
FT	/*lag-a
FT	/product= "fcdh605 protein"
XX	
PN	WO200006206-A1.
XX	
PD	10-FEB-2000.
XX	
PF	30-JUL-1999; 99WO-US17394.
XX	
PR	30-JUL-1998; 98US-0126640.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
P1	Falb DA;
XX	

DR WP1: 2000-205414/18.
DR P-PSDB: Y45017.

Identifying substances for ameliorating symptoms of fibroproliferative diseases or oncogenesis related disorders -

Examples: Fig 5: 214pp: English.

The patent discloses methods for the treatment and diagnosis of cardiovascular diseases by novel human genes which are differentially expressed in different cardiovascular disease states. Compositions which can modify TGF-beta signalling pathway are identified by screening. These are used therapeutically to treat fibroproliferative and oncogenic disorders, especially TGF (Transforming growth factor)-beta related disorders, including diabetic retinopathy, arteriosclerosis, pancreatic cancer, angiogenesis, inflammation, fibrosis, tumour growth and vasculature. The present sequence is fchd605 gene which is up-regulated in monocytes treated with oxidised LDL (low density lipoprotein) can be used to design cardiovascular disease treatment strategies. Depending on whether the up-regulation has a pathogenic or protective effect treatment methods can be designed to increase or decrease the activity of the protein product of the gene.

Sequence 1228 BP: 265 A: 361 C: 341 G: 261 T: 0 other:

Query Match 100.0% Score 1228; DB 21; Length 1228;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATCTGTCATCTGCGACGCTGCCAGCCGACCATGCTGAGCCCGCCGCGCCGCGC 60
DB 1 ATCTGTCATCTGCGACGCTGCCAGCCGACCATGCTGAGCCCGCCGCGCCGCGC 60
OY 61 CCTTCACCATCCCGGAGCCCGGCGGCGCTCGGTCTGAGATCTTCACTTCGACCT 120
DB 61 CCTTCACCATCCCGGAGCCCGGCGGCGCTCGGTCTGAGATCTTCACTTCGACCT 120
OY 121 CTCTCCGAGACCGGACGCGCCCTCGCGCCGCGCCGCGCTCTGCGGCGACCGAAG 180
DB 121 CTCTCCGAGACCGGACGCGCCCTCGCGCCGCGCCGCGCTCTGCGGCGACCGAAG 180
OY 121 CTCTCCGAGACCGGACGCGCCCTCGCGCCGCGCCGCGCTCTGCGGCGACCGAAG 180
DB 121 CTCTCCGAGACCGGACGCGCCCTCGCGCCGCGCCGCGCTCTGCGGCGACCGAAG 180
OY 181 CGGAGCGGACGGTCTTACCTCGAGTGTGCTCGGCGCGCAGCTGCGCATCGACGACCG 240
DB 181 CGGAGCGGACGGTCTTACCTCGAGTGTGCTCGGCGCGCAGCTGCGCATCGACGACCG 240
OY 181 CGGAGCGGACGGTCTTACCTCGAGTGTGCTCGGCGCGCAGCTGCGCATCGACGACCG 240
DB 181 CGGAGCGGACGGTCTTACCTCGAGTGTGCTCGGCGCGCAGCTGCGCATCGACGACCG 240
OY 241 AACCCAGCCAAAAGGCTTCTTCTTCTGCTGCTACCATGCTTCTTGCAGATCTGATG 300
DB 241 AACCCAGCCAAAAGGCTTCTTCTTCTTCTGCTGCTACCATGCTTCTTGCAGATCTGATG 300
OY 301 GCTGAAGAGGCTGTGCGCGCGCCCTGCGCTCCAGAGAGCGCCCTTAACGCGCATGCTG 360
DB 301 GCTGAAGAGGCTGTGCGCGCGCCCTGCGCTCCAGAGAGCGCCCTTAACGCGCATGCTG 360
OY 361 GCGCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 361 GCGCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
OY 421 TAGCTCTGAGCTTCAGACATTTCTCTCAGACAAACACCGCGCGCTTCTTAAGTACTG 480
DB 421 TAGCTCTGAGCTTCAGACATTTCTCTCAGACAAACACCGCGCGCTTCTTAAGTACTG 480
OY 481 CCGGCAATCTCCGCAAAAAGATCGAANAACCAAGAAACACGAGCGATGCTGCTGCG 540
DB 481 CCGGCAATCTCCGCAAAAAGATCGAANAACCAAGAAACACGAGCGATGCTGCTGCG 540
OY 541 CGAAGCGCTATCCCACTGAGCTTCCGAGCGCACTTTCGAGCTCAGACATTCAGACCGCA 600
DB 541 CGAAGCGCTATCCCACTGAGCTTCCGAGCGCACTTTCGAGCTCAGACATTCAGACCGCA 600
OY 601 GAGCGCACCGGCTGCTGAGCGCGGAGCGGAGCGGCTGACAGAGACGCGCGCATAGAGACG 660
DB 601 GAGCGCACCGGCTGCTGAGCGCGGAGCGGAGCGGCTGACAGAGACGCGCGCATAGAGACG 660

```

```

OY 661 GAGCGCACCGGCTGCTGAGCGCGGAGCGGAGCGGCTGACAGAGACGCGCGCATAGAGACG 720
DB 661 GAGCGCACCGGCTGCTGAGCGCGGAGCGGAGCGGCTGACAGAGACGCGCGCATAGAGACG 720
OY 721 TATTGGCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
DB 721 TATTGGCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
OY 781 GTACGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
DB 781 GTACGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
OY 841 AGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 AGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
OY 901 GCGCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 GCGCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
OY 961 AAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
DB 961 AAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
OY 1021 TCTGCTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
DB 1021 TCTGCTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
OY 1081 TCTGCTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
DB 1081 TCTGCTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
OY 1141 CAGCGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 CAGCGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
OY 1201 TGTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1260
DB 1201 TGTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1260

```

RESULT 2
T94471
ID T94471 standard: cDNA: 1228 BP.
XX T94471:
XX
XX 03-MAR-1998 (first entry)
XX
XX Human Fchd605 gene differentially regulated in monocytes.
XX
XX Fchd605 gene: differential expression; monocyte; human;
XX foam cell; cardiovascular disease; arteriosclerosis; ischemia;
XX reperfusion; hypertension; restenosis; arterial inflammation;
XX therapy; diagnosis; drug screening; marker; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 1..471
XX CDS /*tag= a
XX
XX MO9730065-A1.
XX
XX 21-AUG-1997.
XX
XX 14-FEB-1997: 97WO-US02291.
XX
XX 13-FEB-1997: 97US-0799910.
XX 16-FEB-1996: 96US-0011787.
XX
XX (MILL-) MILLENNIUM PHARM INC.

XX Falb DA:
 PI WPI: 1997-424966/39.
 DR P-PSDB: W36006.

XX New genes differentially expressed in cardiovascular disease - used
 PT for diagnosis, drug screening and treatment of cardiovascular
 PT disease, e.g. atherosclerosis, restenosis, hypertension, etc

XX Claim 1: Fig 5; 163pp; English.

XX Fchd605 is a novel human gene that is up-regulated in monocytes
 CC treated with oxidised low density lipoproteins (LDL). Differential
 CC display was used to detect genes that are differentially expressed
 CC in monocytes treated so as to simulate the conditions under which
 CC foam cells develop during atherogenesis. Both fchd605 and fchd602
 CC (see T94470) are up-regulated under the disease condition of
 CC treatment with oxidised LDL. The fchd605 gene product (see W36006)
 CC has sequence similarity to mouse gly96. The discovery of the
 CC up-regulation of these 2 genes provides a fingerprint profile,
 CC e.g. markers for the study of cardiovascular diseases, including
 CC atherosclerosis, ischemia/reperfusion, hypertension, restenosis,
 CC and arterial inflammation. Methods are provided for the diagnosis,
 CC monitoring in clinical trials, screening for therapeutically
 CC effective compounds, and treatment of cardiovascular diseases based
 CC on discoveries regarding the expression patterns of novel genes
 CC fchd531 (see T94467), fchd540 (see T94468), fchd545 (see T94469),
 CC fchd602 and fchd605.

XX Sequence 1228 BP; 265 A; 362 C; 340 G; 261 T; 0 other:

Query Match 99.98; Score 1226.4; DB 18; Length 1228;
 Best local Similarity 99.98; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;
 Matches 1227; Conservative 0;

QY 1 ATGTGTCATCTCGACGCTGCACCGACCATGACCATCTCTGAGCGCCCGACCGGCC 60
 DB 1 atgtgtcatctctcgacgtctgcacccgacatcctctcgagcccgaccccgcc 60
 QY 61 CGCTCCACATCCCGGACCCCGGGGGCTCCGGTCTGAGTCTTTCACCTTGACCT 120
 DB 61 cccctccacatcccggaaccccggggtccggtcgtgagatctctacatcttcgacct 120
 QY 121 CTCCCGAGCGCCGACGGCGCCGCGCGCGCCGACGCGCTCTGCGCGACCGCAAG 180
 DB 121 ctcccgagccgcgaagcgcccgcgcgcccgcccgcccgcccgcgcgagcgaaag 180
 QY 181 CGAGCGCGACGGGTTCTACCTCGATGCTCGGCGCGACGCTGCAGTGAAGAACG 240
 DB 181 cgcagcgcgaggggttctctactcctcgagtgctcgcgccagctgcagagcgagaa 240
 QY 241 AACCCAGCCAAAGGCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
 DB 241 aaccacagccaaaggtct 300
 QY 301 GCTGAAGAGGCTGAGCGCGCGCCCTGCTCAAGAGACGCCCTTAAGCGCCATCCCT 360
 DB 301 gctgaagaggtgtagcgcgccctctctctctctctctctctctctctctctctctct 360
 QY 361 GCGGCCAACCCTGTGTCTCCCGCTCTGAGCCCTTTAATTGACTTGGAGCCCTG 420
 DB 361 gcgcccaaccctgtgtctcccgctctgagcccttttaattgacttggagcccttgag 420
 QY 421 TAGGCTGTGAGCTCAGACACTTCTCTCAGCAACACCGCGGCTCTTAAGTGTGACT 480
 DB 421 taggctgtgagctcagacacttctctcagcaacacccgcgctcttaagtgtgactc 480
 QY 481 CCCCCACTTCCCAAAAGAAATCCGAAACACCAAGAAACACGAGCGCTGATCTGTC 540
 DB 481 cccgaccttcccaaaagaatccgaaacaccaaagaagaacacgagcgatcctgltcg 540

QY 541 CGAGGCGTATCCCAATCCGACCTTCGAGGCACTTGACTGAGACACTACACGCA 600
 DB 541 cgaagcgatctcccaactcggacttcgagcaacttgacacacacacacgagcgga 600
 QY 601 GAGCCACCCGGTCTTGTAGGCGCGGACCGGACGACGACGACGACGACGACGAC 660
 DB 601 gacccacccgggtcttgaagcgagcgagcgacgaagacgagcgagcgacgaagac 660
 QY 661 GAGCCACGCGACCTGCGGCTAGCGCCGTGGGAAGACAGACGCTGTTAATTATTT 720
 DB 661 gaggcacagcccgactcgggctagggccgggtgggaaggaagagtgcttaatttct 720
 QY 721 TATTGCTCTAATTATTAATTATTAATTATTAATTATTAATTATTAATTATTA 780
 DB 721 tattgctctaatattataattataattataattataattataattataattataatt 780
 QY 781 GTACTGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 840
 DB 781 gtactgatatattatattatattatattatattatattatattatattatattat 840
 QY 841 AGGTCCTTGTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 900
 DB 841 aggtccttgatattatattatattatattatattatattatattatattatattat 900
 QY 901 GCGAAGTAGGAGAAAGAAATGGGAGACTGCGTGGCGAGACGCTCCGGCTGGGAT 960
 DB 901 ggcgaagtaggagaaagaaatgggagactgcgaggcgagacgctccggctgggatg 960
 QY 961 AAGTCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 DB 961 aagtcgtgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1020
 QY 1021 TCTGTCTCTGT 1080
 DB 1021 tctgtctctgt 1080
 QY 1081 TCTTTTAAAGTCCGCTTTAGGCTGCTGCGAGGATGAGAGGTTGGGTTGGTGT 1140
 DB 1081 tctttttaaagtccgctttaggctgctgcgaggatgagaggttgggttgggtgtgt 1140
 QY 1141 CACGAGCGACCTGTGATGATGCGCTGATGATGCTGCTGCTGCTGCTGCTGCTG 1200
 DB 1141 cacgagcgacctgtgatgatgctgctgctgctgctgctgctgctgctgctgctg 1200
 QY 1201 TGTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1260
 DB 1201 tgtcaaaaaaataaataaataaataaataaataaataaataaataaataaataa 1260

RESULT 3
 280483/C
 1D 280483 standard; cDNA; 427 BP.
 XX 280483:
 AC 07-APR-2000 (first entry)
 XX
 DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:567.
 XX
 KW Human: gene expression product; diagnosis; tumour; colon cancer;
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
 KW cytosolic; sarcoma; breast cancer; neoplasia; dysplasia;
 KW hyperplasia; ds.
 XX
 OS Homo sapiens.
 XX
 PN W09964576-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 09-JUN-1999; 99WO-1B01062.
 XX
 PR 10-JUN-1998; 98US-0088801.

CC the polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 1126 BP: 101 A; 102 C; 575 G; 57 T; 291 other;
SQ

SQ Sequence 1126 BP; 101 A; 102 C; 575 G; 57 T; 291 other;

```

Query Match      4.0%; Score 49.6; DB 21; Length 1126;
Best Local Similarity 33.2%; Pred. NO. 0.0012;
Matches 192; Conservative % 0; Mismatches 383; Indels 4; Gaps 1;

```

[illegible]

RESULT 12

T20789
ID T20789 standard: cDNA to mRNA: 53 bp.

XX T20789;

XX
DT 09-JUL-1996 (first entry)

XX
DE Human gene signature HUMGS02013.

XA Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.

XX	Homo sapiens.
OS	

XX	WO9514772-A1.	
PN		
XX		
PD	01-JUN-1995.	
XX		
PF	11-NOV-1994;	94WO-JP
XX		
PR	12-NOV-1993;	93JP-03
XX		
PA	(MATS/) MATSUBARA K.	
PA	(OKUB/) OKUBO K.	
XX		
PI	Matsubara K,	Okubo K;
XX		
DR	WPI: 1995-206931/27.	

xx Identifying gene signatures in 3'-directed human cDNA library - e.g.
 xx PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues

PS Claim 1; Page 723; 2245pp; Japanese.

A single-stranded DNA (or its complementary strand or the corresp. double-stranded DNA) which comprises one of the 7837 "GS" sequences given in T91001-T926837 and which is able to hybridise to part of human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were obtained from 3'-directed cDNA libraries prepared from various human tissues; synthesis of cDNA was initiated from the 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-untranslated sequence is unique to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library is constructed so as to reflect accurately the relative abundance of different mRNAs in the particular tissue from which it was derived. The appearance frequency of a given GS in a cDNA library can be determined (esp. using primers and probes derived from the GS sequences) as a means of diagnosing abnormal cell function or recognising different cell types.

Sequence 53 BP; 18 A; 9 C; 9 G; 17 T; 0 other;
XX

Query Match	3.9%	Score 48.2;	DB 16;	Length 53;
Best Local Similarity	94.3%	Pred. No. 0.00069;		
Matches 50;	Conservative	0;	Mismatches 3;	Indels 0;
			Gaps	0;

Qy	1158	GATCGCCTAGTATGTTCTGTGAACACAATAAAATTGATTACTGTCAAAAAA	1210
Db	1	gatcgctagtattttctgtgaacacacaataaaattgatctactgtctgcaga	53

RESULT 13
251683

ID	251683 standard; cDNA; 4228 BP.
1	251683 standard; cDNA; 4228 BP.

AC 251683:

XX
DT 04-JUL-2000 (first entry)XX
DE Human cyclic nucleotide-associated protein-2 (CNAP-2) cDNA.

Cyclic nucleotide-associated protein-2; CNAP-2; human; cytostatic;
anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;
immunomodulatory; anti-asthmatic; anti-anaemic; anti-diabetic; diagnosis;
anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;
anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;
anti-infertility; anti-allergic; vasotropic; immunosuppressive;
hypotensive; gene therapy; prevention; treatment; arteriosclerosis;
cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;
neurological; vision; reproductive; smooth muscle; ss.
Homo sapiens.

OS Homo sapiens.

XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
PA
XX
PI Williams LT, Escobedo
PI Reinhard C, Giese K
PI Lamson G, Drmanac R
PI Leshkowitz D, Kita
XX WPI: 2000-126369/11.

cells -
Polynucleotide library used to deter-
mine -
Claim 1; Page 1004; 1097pp; English.

```

XX  A00010 to A02716 represent polynucleotides isolated from cDNA libraries
CC  constructed from human colon cancer cell lines. The present invention
CC  also describes a method of detecting differentially expressed genes
CC  correlated with a cancerous state of a mammalian cell, comprising
CC  detecting at least one differentially expressed gene product in a test
CC  sample derived from a cell suspected of being cancerous, where detection
CC  of the differentially expressed gene product is correlated with a
CC  cancerous state of the cell from which the test sample was derived.
CC  The polynucleotides sequences can be used in a method for detecting
CC  differentially expressed genes correlated with a cancerous state of a
CC  mammalian cell. The polynucleotides can also be used as probes for
CC  detecting and mapping related genes. They can be used in diagnosis and
CC  prognosis of diseases and disorders (e.g. identification of
CC  pre-metastatic or metastatic cancerous states, stages of cancer, or
CC  responsiveness of cancer to therapy). This is particularly for breast
CC  cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC  negative breast cancer, lung cancer, and colon cancer.
XX
SO  Sequence 1593 Bp; 133 A; 49 C; 808 G; 49 T; 554 other:

      Query Match      3.6%; Score 44.8; DB 21; Length 1593;
      Best Local Similarity 32.9%; Pred. No. 0.025;
      Matches 155; Conservative 0; Mismatches 314; Indels 2; Gaps

          22  CACCCCGACCATGACCATCTGCGAGCCCGAGCCCGCCGCGCCCTCCACCATCCCGGGACCC 81
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy	82	CGCGGGGCTCCGGTCTTGAGATCTTCACCTTCGACCCTCCTCCGGAGCGCGCAGCGGCC	141
Db	958	NNNCNCTNTTCNTNNCNNNNNGCCCCCNCCNNCNCNNNNNNCCC--CCAACNT	900
Qy	142	CCTGCGGGGCGCCCAAGCGCTCTCGGGGCACCGAAGCAGCGAGCGCGGGGTCTCTTAC	201
Db	899	NNGNCTTNNNCCCCCCCCCNCCNNGCNCNGNCCNCCNCCNCCNCCNCCNCCCTCC	840
Qy	202	CCTCGAGTGGTCCGGCGCCAGCTGCGCAGTCGAGGAACCGAACCCAGCGCAAAAGGCTTCTC	261
Db	839	CCCCCNTNNCNCNNNTNNCNCNNCNGNNCNCNCCNCCNCCNCCNANNNNCNCNC	780
Qy	262	TTTCTGCTGTCACCATGCTCTCTGCCAGATCTTGATGGCTGAAGAGGGTGTGCGCGCG	321
Db	779	GCCCCCCCCCNCCNNCCNTCCCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCCTTCC	720
Qy	322	CCCTGCTCTCACAGAGACGCGCCCTAACGCGCATCTCTGGCGCCACGCCCTGTGTCCCCC	381
Db	719	CCCCNTCCCCCCCCNTCCCNNTTCGCNCTCTNCCCNCCNCCCNACCNTNNCCCNCC	660
Qy	382	GTCTCTGAGCGCTTTAATCTGACTTCGGAGCGCTCGGACTACGCTCTGAGCTTCAGCACT	441
Db	659	C-NCTNNCCCCGCGNTTCCGNTWTCCNCNTTCCCTTCCNCCNCCNCCNCCNCCNCCNCC	601
Qy	442	TTCTTCAGGAACAACCGGGCGGCTTCTTACTGTGACTTCCCCCGCACTCCCC	492
Db	600	TNNCCNCCNCCNCCANCCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC	550

Search completed: January 30, 2001, 21:29:29
Job time: 19255 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:11:29 ; Search time 64.84 Seconds
(without alignments)
155.458 Million cell updates/sec

Title: US-08-799-910-10_COPY_71_156

Perfect score: 442

Sequence: 1 VRQLPVEPNPAKRLFL.....EPSDYALDLSTFLOQHPAAF 86

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_15:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	442	100.0	193	4	075353
2	84.5	19.1	637	4	094878
3	81.5	18.4	102	5	09T2D3
4	81.5	18.4	452	5	09UAS5
5	81.5	18.4	468	5	09XY65
6	80.5	18.2	302	12	09QM69
7	78	17.6	4957	4	014687
8	78	17.6	5262	4	014686
9	77	17.4	1568	4	095785
10	77	17.4	2424	5	09VZ48
11	76	17.2	148	2	09JZ87
12	76	17.2	801	5	09W3K6
13	76	17.2	1677	5	000805
14	74	16.7	114	12	09VYX8
15	74	16.7	955	11	088287
16	74	16.7	1040	5	09VH04
17	74	16.7	1561	11	088286
18	73.5	16.6	925	12	091338
19	73	16.5	1029	12	09YMX0

20	72	16.3	659	5	09VQ23
21	72	16.3	852	4	09NP71
22	71	16.1	359	5	09TYW7
23	71	16.1	1882	3	09P898
24	70.5	16.0	244	12	069126
25	70.5	16.0	1900	4	015022
26	70.5	16.0	2605	4	09UK25
27	70.5	16.0	2715	4	09UMN6
28	70	15.8	148	2	09JUV2
29	70	15.8	1233	2	09R132
30	70	15.8	2971	4	09Y5L9
31	69.5	15.7	277	4	09NXN6
32	69.5	15.7	562	4	09NVJ4
33	69.5	15.7	745	4	09P2N8
34	69.5	15.7	757	6	09N015
35	69.5	15.7	831	4	09UMS6
36	69.5	15.7	1887	3	012533
37	69	15.6	418	5	018503
38	69	15.6	646	4	09ULX6
39	69	15.6	646	4	09UGM0
40	69	15.6	1454	5	010463
41	69	15.6	2157	11	09Z1R1
42	69	15.6	2187	11	P70670
43	68.5	15.5	701	4	09NXX8
44	68.5	15.5	734	4	09NSE6
45	68.5	15.5	821	4	09UII8

ALIGNMENTS

RESULT 1

075353
10 075353 PRELIMINARY; PRT: 193 AA.
AC 075353;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10; Last annotation update)
DE ANTI-DEATH PROTEIN.
GN IEX-1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=98369175; PubMed=9703517;
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival.";
RL Science 281:998-1001(1998).
DR EMBL; AF039067; AAC32558.1; -;
DR EMBL; AF071596; AAC72344.1; -;
SQ SEQUENCE 193 AA; 21028 MW; 7927D9D3FFBC7C57 CRC64;

Query Match 100.0%; Score 442; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 8.7e-40;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRQLPVEPNPAKRLFLLLTIVFCQILMAEGVPAPLPPEDPAPNAASLAPTVPSPVLE 60
|||||
Db 108 VRQLPVEPNPAKRLFLLLTIVFCQILMAEGVPAPLPPEDPAPNAASLAPTVPSPVLE 167
|||||

QY 61 PFNLTSPSPDYALDLSTFLOQHPAAF 86
|||||

Db 168 PFNLTSPSPDYALDLSTFLOQHPAAF 193
|||||

RESULT 2

094878
ID 094878 PRELIMINARY; PRT: 637 AA.
AC 094878;

DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE KIAA0781 PROTEIN (FRAGMENT).
 GN KIAA0781.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
 KA Kotani H., Nomura N., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 5:277-286(1998).
 DR ENBL; AB018324; BAA34501.1; .
 DR INTERPRO; IPR000449; .
 DR INTERPRO; IPR002965; .
 DR PRINTS; PR01217; PRICHEXTENSN.
 FT NON_TER
 FT I
 SQ SEQUENCE 637 AA: 70516 MW: DEBF005BA9B9CDD CRC64;

Query Match 19.1%; Score 84.5; DB 4; Length 637;
 Best Local Similarity 33.3%; Pred. No. 0.41;
 Matches 30; Conservative 12; Mismatches 27; Indels 21; Gaps 5;

OY 7 VEEPNAKRLFL-----LTIIVFCOILMAEGVP-----APLPEDAPNAASLAP----- 52
 Db 423 LOEHRIQKRLFLQKOSLOAQFNQMTAESYPOPSQOLPLPROETPPPSQOAPFSLT 482

OY 53 TPVSPVLEPNTSEPSDYALDLSTFLQOH 82
 Db 483 QPLSPVLEP-----SSEQMQY-----SPFLSQY 505

RESULT 3
 OYTD3
 ID OYTD3 PRELIMINARY; PRT; 102 AA.
 AC OYTD3
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE R160.5 PROTEIN.
 GN R160.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Petoderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RL Nature 368:32-38(1994).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;

RA Nelson J., Gattung S.;
 RT "The sequence of C. elegans cosmid R160."
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF099001; AAC68734.1; .
 SQ SEQUENCE 102 AA: 11426 MW: AFD366C3932EB26C CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 102;
 Best Local Similarity 24.7%; Pred. No. 0.13;
 Matches 19; Conservative 19; Mismatches 28; Indels 11; Gaps 3;

OY 11 NPAKRLFL-----LTIIVFCOILMAEGVPAPLPEDAPNAASLAPPPSVLEPFN-- 63
 Db 7 NPAIQFAEVVILVCAVCICMLIERADGFCPLPPEELRRIVNTIP-PRTPIAAPVNR 65

OY 64 ---LTSEPSDYALDLST 77
 Db 66 QVVIKIDPTESDVDMTS 82

RESULT 4
 OYUAS5
 ID OYUAS5 PRELIMINARY; PRT; 452 AA.
 AC OYUAS5
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE DEVELOPMENTAL-SPECIFIC PROTEIN CONZAS8 (FRAGMENT).
 GN CONZAS8.
 OS Euplotes crassus.
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;
 OC Euplotes.
 OX NCBI_TaxID=5936;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ST9, ST11;
 RX MEDLINE=97206572; PubMed=9172827;
 RA Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;
 RT "Conjugation-specific genes in the ciliate Euplotes crassus: gene
 RT expression from the old macronucleus".
 RL J. Eukaryot. Microbiol. 44:1-11(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ST9, ST11;
 RA Jacobs M.E., Ling Z., Klobutcher L.A.;
 RT "conzA8 encodes a novel and abundant protein targeted to the
 RT developing macronucleus in Euplotes crassus".
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF116196; AAD29625.1; .
 FT NON_TER
 FT I
 SQ SEQUENCE 452 AA: 49763 MW: ASC13C49ECA998EA CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 452;
 Best Local Similarity 34.5%; Pred. No. 0.61;
 Matches 30; Conservative 10; Mismatches 30; Indels 17; Gaps 5;

OY 3 RQLPVEEP-----NPAKRLFLLLTIVFCOILMAEGVPAPLPEDAPNAASLAPT-----p 54
 Db 316 RSQPTPLPDILENPAKEIASSLSKMI---QETAIESITPQPP-----VSSIQPSAPVHP 367

OY 55 VSPVLEPFN-LTSEPSDYALDLSTFLQ 80
 Db 368 SAPSLEPSSTLASEPRDIIPDSAAALK 394

RESULT 5
 OYX65

XX	MEDLINE=97388474; PubMed=9247308;
RA	Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA	Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA	Canaani E.;
RT	"Structure and expression pattern of human ALR, a novel gene with
RT	strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT	trithorax".
RL	Oncogene 15:549-560(1997).
DR	ENBL: AF010403; AAC51734.1; -.
DR	INTERPRO: IPR001214; -.
DR	INTERPRO: IPR001822; -.
DR	INTERPRO: IPR001841; -.
DR	INTERPRO: IPR001965; -.
DR	INTERPRO: IPR002965; -.
DR	PFAM: PF00628; PBD: 5.
DR	PFAM: PF00856; SET: 1.
DR	PRINTS: PRO1217; PRICHTENS.
DR	PROSITE: PS00398; RECOMBINASES_2; UNKNOWN_1.
DR	SEQUENCE 5262 AA: 564181 MW: 26B7C74CA0417E44 CRC64

Query Match 17.6%; Score 78; DB 4; Length 5262;
Best Local Similarity 30.4%; Pred. No. 18;
Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

Qy	1	VRRLPVEEPNDPAKRLFLLLTIVFCQILMAEEGVAPILPPEDAPNAASLAPTPVSPVLE	60
		: : : : : : : : :	
Db	4604	LKQESAPEPTQHRHYNVNSLNDVRQL-----SAPPEEPSPPSPILAPSPAPPT	4656

Qy 61 PF-NLTSEP 68

Db 4657 PLVELPTEP 4665

RESULT	9
ID	O95785
AC	PRELIMINARY; PRT; 1568 AA.
DT	01-MAY-1999 (Tremblrel. 10, Created)
DT	01-WAY-1999 (Tremblrel. 10, Last sequence update)
DE	01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE	HUMAN HOMOLOG OF MUS MUSCULUS WIZL PROTEIN (HUMAN HOMOLOG OF MUS
DE	MUSCULUS WIZS PROTEIN) (FRAGMENT)".
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI	TaxID:9606.

[1]
 SEQUENCE FROM N.A.
 Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
 Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
 Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 Dangnan L., Erier A., Christensen M., Georgescu A., Avila J., Liu S.,
 Attix C., Andreise T., Frankheim M., Amico-Keller G., Coefield J.,
 Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
 Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 Olsen A.S., Carrano A.V.:
 "Sequence analysis of a 3.5 Mb region in 19p13.1 between OLFR and
 JUND".
 Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

[2] SEQUENCE OF 783-1568 FROM N.A.
Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
Burkhardt-Schultz K., Gordon L., Dias J., Brower A., Ramirez M.,
Stilwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,
Carnes J., Danganan L., Erler A., Christensen M., Georgescu A.,
Avila J., Liu S., Attix C., Andreise T., Trankheim M.,
Amico-Keller G., Cosfield J., Duarte S., Lucas S., Bruce R.,
Thomas P., Quan G., Kronmiller B., Arellano A., Sanders C., Ow D.,
Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.:
"sequence analysis of an ~1.5 Mb contig in 19p13.1 between OLF and
D19S485 ";
Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AC007059; AAd19818.1; -
DR EMBL; AC006128; AAC97985.1; -
DR EMBL; AC007059; AAd19817.1; -
DR INTERPRO; IPR000822; -
DR PFAM; PF00096; zf-C2H2; 10.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
KW Zinc-finger; Metal-binding; DNA-binding.
FT NON_TER 1
SQ SEQUENCE 1568 AA; 170077 MW; 74F0D34DA5F565D0 CRC64;

Query Match 17.4%; Score 77; DB 4; length 1568;
Best Local Similarity 40.4%; Pred. No. 6;
Matches 19; Conservative 3; Mismatches 17; Indels 8; Gaps 3

Oy 30 MAEGVPAIPUPPEDAPNAASLAPTPVSPVLE-----PFNLTSEP 68
: : : : : : : : : : : : : : : :
ob 908 LALAGSTPKNPEDKSPQLSLRPPASPKAQWPOSDEGPLNLTSGP 954
: : : : : : : : : : : : : : : :

```

RESULT 10

Q9VZ48
ID Q9VZ48 PRELIMINARY; PRT; 2424 AA.

AC O9VZ48;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE CG2174 PROTEIN.
GN CG2174.
GN

OS *Drosophila melanogaster* (Fruit fly)

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;
BY MEDIAN=2018600; RUN=1072222

RA Adams M.D., Celniker S.E., Holt R.
MEDLINE=20158008; PubMed=10731132;
RX

RA Amanatides P.G., Scherer S.E., Li

RA George R.A., Lewis S.E., Richards
RA Sutton G.G., Wortman J.R., Vandell

RA Brandon R.C., Rogers Y.-H.C., Blaz

RA Wan K.H., Doyle C., Baxter E.G., H
RA April J.F. Achavani A. Anu - 7

RA Ballew R.M., Basu A., Baxendale J.

RA Beeson K.Y., Benos P.V., Berman B.

RA Burtis K.C., Busam D.A., Butler H.
RA Borkov D., Bolchan M.R., Bouck J.

RA Cherry J.M., Cawley S., Dahlke C.,

RA de Pablos B., Belcher A., Deng Z.,
 Dodson K., Doup I., Downes M.

RA Durbin K.J., Evangelista C.C., Fer

RA Foster C., Gabrielian A.E., Garg N
RA Cloddek A
RA Cloddek A

[illegible]

RA Hostin D., Houston K.A., Howland T.

RA Jatali M., Kalush F., Karpen G.H.,
RA Kimmel B.E., Kodjira C D Kraft C

RA Lasko P., Lei Y., Levitsky A.A., L

RA. Liu X., Mattei B., McIntosh T.C., ...
RA. Merkley C. ...

RA Mount S.M.; Moy M.; Murphy B.; Murn

Nelson D.R., Nelson K.A., Nixon K.

RA Palazzolo M., Pittman G.S., Pan S.
RA Reinert K., Remington K., Saunders

RA Shue B.C., Siden-Kiamos I., Simpson

RA Spier E., Spradling A.C., Stapleton
RA Svirskaas B. Testes C. Turner B.

Wang Z.-Y., Wassarman D.A., Weinstein

RA Williams S.M., Woodage T., Worley P.

IE J., YEH R.-F., Zaveri J.S., Zhang
Zheng X.H., Zhong F.N., Zhong W.

Figure 1


```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003484; AAF47980.1; -.
DR HSSP: P08799; 1MD.
DR FLYBASE: FBgn0030252; CG2174.
DR INTERPRO: IPR000048; -.
DR INTERPRO: IPR000857; -.
DR INTERPRO: IPR001609; -.
DR PFAM: PF00063; myosin_head; 1.
DR PFAM: PF00612; 10; 3.
DR PFAM: PF00784; MYTH4; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
SQ SEQUENCE 2424 AA; 267616 MW; 8AD62AA33F9AA5D0 CRC64;

Query Match
Best Local Similarity 17.4%; Score 77; DB 5; Length 2424;
Matches 22; Conservative 7; Mismatches 29; Indels 2; Gaps 2;

QY 28 ILMAEEGVPAPLPEDAPNAASLAPTPVSPVLEPENLTSEPS-DVALDLSTFLOOHPAAF 86
Db 1950 LLPPAPVPAPPPPIRPSMAPPP-APQSPPTARSPEDPRTSSQVVKCHVPAF 2008

RESULT 11
Q9J2K7 PRELIMINARY; PRT; 148 AA.
ID Q9J2K7
AC Q9J2K7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ACYL COA THIOESTER HYDROLASE FAMILY PROTEIN.
GN NMB0925.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=2017555; PubMed=10710307;
RA Tottelien H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathavan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58.";
RL Science 287:1809-1815(2000).
DR EMBL: AE002444; AAF41333.1; -.
DR TIGR: NMB0925; -.
KW Hydrolase.
SQ SEQUENCE 148 AA; 16332 MW; 28E9B40D502D80B4 CRC64;

Query Match
Best Local Similarity 17.2%; Score 76; DB 2; Length 148;
Matches 15; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 18 FULLTIVFCQILMAEEGVPAPLPEDAPNAASLAPTP 54
Db 112 VLVTEAVTYVAIDAGNPRPIPEGNPKLAGLLPTP 148

RESULT 12
Q9W3K6 PRELIMINARY; PRT; 801 AA.
ID Q9W3K6
AC Q9W3K6
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DR CG2258 PROTEIN.
GN CG2258.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,
RA Ballev R.M., Basu A., Raxendale J., Bayraktaroglu L., Heasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bröttier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Hochs S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003442; AAF46320.1; -.
DR HSSP: P06241; 1SHF.
DR FLYBASE: FBgn002997; CG2258.
DR INTERPRO: IPR001452; -.
DR PFAM: PF00018; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
SQ SEQUENCE 801 AA; 89233 MW; A0F5C824663DC391 CRC64;

Query Match
Best Local Similarity 17.2%; Score 76; DB 5; Length 801;
Matches 21; Conservative 8; Mismatches 23; Indels 10; Gaps 2;

QY 26 CQILMAEEGVPAPLPEDAPNAASLAPTPVS-----PVLEPNLTSEPSDYALDL 75
Db 206 CTIOTPOQQQLPAPPAPAPSPSAASATPTAATSKAATCGDVAOUNITSEPTTTT 265

QY 76 ST 77
Db 266 TT 267

RESULT 13

```

Db 70 S 70

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 21:29:34 ; Search time 57.93 seconds
(without alignments)
401.549 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPSYALDSTFLOQHPAAF 156

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 912318 seqs, 149113690 residues

Total number of hits satisfying chosen parameters: 912318

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending_Patents_AA:*
- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
 - 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
 - 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
 - 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
 - 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
 - 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
 - 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
 - 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
 - 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
 - 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
 - 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
 - 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
 - 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
 - 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
 - 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
 - 21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
 - 22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
 - 23: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
 - 24: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
 - 25: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
 - 26: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
 - 27: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
 - 28: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	823	100.0	156	11	US-08-799-910-10
2	823	100.0	156	12	US-08-825-486-10
3	823	100.0	156	12	US-08-826-248-10
4	823	100.0	156	12	US-08-870-434-11
5	823	100.0	156	13	US-08-925-588-10

Sequence 10, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 11, Appl
Sequence 10, Appl

6	823	100.0	156	17	US-09-372-044-10
7	794	96.5	151	22	US-60-197-873-17386
8	440	53.5	123	22	US-60-196-718-6102
9	425	51.6	83	1	PCT-US00-15136-168
10	423	51.4	83	1	PCT-US00-15136-96
11	422	51.3	83	1	PCT-US00-15136-167
12	381	46.3	89	22	US-60-196-718-4945
13	358	43.5	71	1	PCT-US00-15136-169
14	358	43.5	92	22	US-60-196-718-5439
15	340	41.3	72	1	PCT-US00-15136-170
16	273.5	33.2	56	22	US-60-196-718-3936
17	114.5	13.9	1828	22	US-60-188-986-15
18	114.5	13.9	2012	22	US-60-229-524-80
19	114.5	13.9	2972	19	US-09-579-181-2
20	114.5	13.9	3041	28	US-60-258-275-409
21	114.5	13.9	3118	19	US-09-579-181-1
22	112.5	13.7	649	22	US-60-191-637-28141
23	112.5	13.7	649	22	US-60-191-681-22750
24	112.5	13.7	1255	22	US-60-167-217-13778
25	112.5	13.7	1255	22	US-60-173-464-11201
26	112	13.6	747	18	US-09-417-507-31296
27	110	13.4	590	1	PCT-US99-22853B-2801
28	109	13.2	8991	11	US-08-714-741-32
29	108	13.1	271	16	US-09-248-796-21050
30	108	13.1	271	22	US-60-096-409-21050
31	107.5	13.1	1013	22	US-60-191-637-39541
32	107.5	13.1	1013	22	US-60-191-681-30574
33	107	13.0	206	9	US-08-529-055-54
34	106	12.9	1110	22	US-60-243-468-898
35	105.5	12.8	1013	22	US-60-173-464-30071
36	105	12.8	296	22	US-60-173-464-18776
37	105	12.8	296	22	US-60-191-637-22907
38	105	12.8	296	22	US-60-191-681-18056
39	104	12.6	129	9	US-08-529-055-53
40	102	12.4	331	22	US-60-173-464-29642
41	102	12.4	331	22	US-60-191-637-38878
42	102	12.4	331	22	US-60-191-681-30133
43	102	12.4	421	18	US-09-451-320-4244
44	101.5	12.3	193	9	US-08-529-055-49
45	101.5	12.3	741	18	US-09-417-507-37506

ALIGNMENTS

RESULT 1
US-08-799-910-10
; Sequence 10, Application US/08799910
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/799,910
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996

ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-067-999
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-799-910-10

Query Match 100.0%; Score 823; DB 11; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRSGSGPEITFDPLPEPAAAPAGPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRSGSGPEITFDPLPEPAAAPAGPSASRGHRK 60

Qy 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPPLPPEDAPNAASL 120
Db 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPPLPPEDAPNAASL 120

Qy 121 APTVPSPVLEPFLNTSEPSDYALDI STFLQOHPAAF 156
Db 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLQOHPAAF 156

RESULT 2
US-08-825-486-10
Sequence 10, Application US/08825486
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE TREATMENT AND DIAGNOSIS OF
TITLE OF INVENTION: CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,486
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-077-999
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-825-486-10

Query Match 100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRSGSGPEITFDPLPEPAAAPAGPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRSGSGPEITFDPLPEPAAAPAGPSASRGHRK 60

Qy 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPPLPPEDAPNAASL 120
Db 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPPLPPEDAPNAASL 120

Qy 121 APTVPSPVLEPFLNTSEPSDYALDI STFLQOHPAAF 156
Db 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLQOHPAAF 156

RESULT 3
US-08-826-248-10
Sequence 10, Application US/08826248
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE TREATMENT AND DIAGNOSIS OF
TITLE OF INVENTION: CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,248
FILING DATE: 28-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-079-999
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein

```
; FRAGMENT TYPE: internal
US-08-826-248-10

Query Match 100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120
Db 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120

Qy 121 APTVPSPVLEPFLNTSEPSDYALDLSLFLQOHPAAF 156
Db 121 APTVPSPVLEPFLNTSEPSDYALDLSLFLQOHPAAF 156

RESULT 5
US-08-925-588-10
; Sequence 10, Application US/08925588
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,588
; FILING DATE: 08-Sep-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-08-925-588-10

Query Match 100.0%; Score 823; DB 13; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120
Db 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120

Qy 121 APTVPSPVLEPFLNTSEPSDYALDLSLFLQOHPAAF 156
Db 121 APTVPSPVLEPFLNTSEPSDYALDLSLFLQOHPAAF 156

; FRAGMENT TYPE: internal
US-08-870-434-11

Query Match 100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120
Db 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120

Qy 121 APTVPSPVLEPFLNTSEPSDYALDLSLFLQOHPAAF 156
Db 121 APTVPSPVLEPFLNTSEPSDYALDLSLFLQOHPAAF 156

RESULT 4
US-08-870-434-11
; Sequence 11, Application US/08870434
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,434
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-084
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-870-434-11

Query Match 100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
```

```
RESULT 6
US-09-372-044-10
; Sequence 10, Application US/09372044A
; GENERAL INFORMATION:
; APPLICANT: Dean FALB et al.
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Treatment and Diagnosis of Cardiovascular Disease
; FILE REFERENCE: 7853-152
; CURRENT APPLICATION NUMBER: US/09/372,044A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-372-044-10

Query Match      100.08; Score 823; DB 17; Length 156;
Best Local Similarity 100.08; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTAPSTIPGRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILQAPTAPSTIPGRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPNTSEPSDYALDLSTFLQHPAAF 156
DB 121 APTVPSPVLEPNTSEPSDYALDLSTFLQHPAAF 156

RESULT 7
US-09-197-873-17386
; Sequence 17386, Application US/60197873
; GENERAL INFORMATION:
; APPLICANT: Bejania, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81.USL PRO
; CURRENT APPLICATION NUMBER: US/60/197,873
; CURRENT FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 17386
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-197-873-17386

Query Match      96.58; Score 794; DB 22; Length 151;
Best Local Similarity 100.0%; Pred. No. 7.5e-58;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTAPSTIPGRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILQAPTAPSTIPGRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPNTSEPSDYALDLSTFLQ 151
```

```
DB 121 APTVPSPVLEPNTSEPSDYALDLSTFLQ 151

RESULT 8
US-60-196-718-6102
; Sequence 6102, Application US/60196718
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AN
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO00456
; CURRENT APPLICATION NUMBER: US/60/196,718
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6102
; LENGTH: 123
; TYPE: PRT
; ORGANISM: HUMAN
US-60-196-718-6102

Query Match      53.5%; Score 440; DB 22; Length 123;
Best Local Similarity 93.5%; Pred. No. 7.4e-29;
Matches 86; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 65 VLYPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASLPTP 124
DB 32 LFLCLPRLRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASLPTP 91

QY 125 VSPVLEPNTSEPSDYALDLSTFLQHPAAF 156
DB 92 VSPVLEPNTSEPSDYALDLSTFLQHPAAF 123

RESULT 9
PCT-US00-15136-168
; Sequence 168, Application PC/TUS0015136
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PS551PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 168
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-15136-168

Query Match      51.6%; Score 425; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 8.2e-28;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 VVRRLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASLPTP 129
DB 1 VVRRLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASLPTP 129

QY 130 EPFNTSEPSDYALDLSTFLQHP 152
DB 61 EPFNTSEPSDYALDLSTFLQHP 83

RESULT 10
PCT-US00-15136-96
; Sequence 96, Application PC/TUS0015136
```

GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PSS51PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 96
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (83)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
PCT-US00-15136-96

Query Match 51.4%; Score 423; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 1.2e-27;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 LPVEEPNPAKRLFLLLTIVFCOILMAEEGVPAIPPPEDAPNAASLAPTPVSPVLEPNL 134
Db 1 LPVEEPNPAKRLFLLLTIVFCOILMAEEGVPAIPPPEDAPNAASLAPTPVSPVLEPNL 60
Qy 135 TSEPSDYALDLSTFLQHPAAF 156
Db 61 TSEPSDYALDLSTFLQHPAAF 82

RESULT 11
PCT-US00-15136-167
; Sequence 167, Application PC/TUS0015136
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PSS51PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 167
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-15136-167

Query Match 51.3%; Score 422; DB 1; Length 83;
Best Local Similarity 98.8%; Pred. No. 1.4e-27;
Matches 82; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 70 VYRROLPVVEEPNPAKRLFLLLTIVFCOILMAEEGVPAIPPPEDAPNAASLAPTPVSPVL 129
Db 1 LVYRROLPVVEEPNPAKRLFLLLTIVFCOILMAEEGVPAIPPPEDAPNAASLAPTPVSPVL 60
Qy 130 EPFNLTSEPSDYALDLSTFLQOH 152
Db 61 EPFNLTSEPSDYALDLSTFLQOH 83

RESULT 12
US-60-196-718-4945
; Sequence 4945, Application US/60196718
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AT
; APPLICANT: USSES THEREOF
; FILE REFERENCE: CL000456
; CURRENT APPLICATION NUMBER: US/60/196,718
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4945
; LENGTH: 89
; TYPE: PRT
; ORGANISM: HUMAN
; US-60-196-718-4945

Query Match 46.3%; Score 381; DB 22; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.7e-24;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRSCHPTMTILOAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRSCHPTMTILOAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
Qy 61 RSRRLVLYPRV 70
Db 61 RSRRLVLYPRV 70

RESULT 13
PCT-US00-15136-169
; Sequence 169, Application PC/TUS0015136
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PSS51PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 169
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-15136-169

Query Match 43.5%; Score 358; DB 1; Length 71;
Best Local Similarity 97.1%; Pred. No. 2.2e-22;
Matches 66; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MCHSRSCHPTMTILOAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRSCHPTMTILOAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
Qy 61 RSRRLVLYP 68
Db 61 RSRRLVLYP 68

RESULT 14
US-60-196-718-5439
; Sequence 5439, Application US/60196718
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AT
; TITLE OF INVENTION: USSES THEREOF
; FILE REFERENCE: CL000456
; CURRENT APPLICATION NUMBER: US/60/196,718
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSEQ for Windows Version 4.0

Search completed: January 30, 2001, 23:10:11
Job time: 6037 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: January 30, 2001, 19:24:41 ; Search time 36.91 Seconds
(without alignments)
75.895 Million cell updates/sec

Title: us-08-799-910-10
Perfect score: 823
Sequence: 1 MCHRSRCHPTMTILOAPTPA.....EPSDYALDLSTFLOQHPAAF 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	823	100.0	156	3	US-08-826-246-10
2	823	100.0	156	3	US-08-944-495-10
3	823	100.0	156	3	US-09-126-640-11
4	95.5	11.6	739	3	US-09-035-648-24
5	95.5	11.6	1596	3	US-09-356-952-3
6	93.5	11.4	905	2	US-08-574-959A-9
7	93.5	11.4	1135	2	US-08-574-959A-7
8	92	11.2	1400	1	US-08-080-255-7
9	92	11.2	1400	3	US-08-465-713-7
10	92	11.2	1400	4	PCT-US93-05857-7
11	90.5	11.0	214	1	US-08-217-327-4
12	88	10.7	468	2	US-08-390-000A-7
13	88	10.7	477	1	US-08-444-734A-4
14	88	10.7	477	1	US-08-087-772A-16
15	87	10.6	2414	1	US-08-227-536-2
16	87	10.6	2414	1	PCT-US95-04682-2
17	85.5	10.4	610	1	US-07-821-717B-6
18	85.5	10.4	610	1	US-08-119-262B-6
19	85.5	10.4	610	1	US-08-135-929A-11
20	85.5	10.4	610	1	US-08-234-265A-11
21	85.5	10.4	774	2	US-08-231-193A-42
22	85.5	10.4	774	2	US-08-486-273A-42
23	85.5	10.4	774	3	US-08-480-474-42
24	85.5	10.4	774	3	US-08-940-086A-42
25	85.5	10.4	1214	2	US-08-231-193A-54
26	85.5	10.4	1214	2	US-08-486-273A-54
27	85.5	10.4	1214	3	US-08-480-474-54
28	85.5	10.4	1214	3	US-08-940-086A-54

29	85.5	10.4	1219	2	US-08-231-193A-50	Sequence 50, Appl
30	85.5	10.4	1219	2	US-08-486-273A-50	Sequence 50, Appl
31	85.5	10.4	1219	3	US-08-480-474-50	Sequence 50, Appl
32	85.5	10.4	1219	3	US-08-940-086A-50	Sequence 50, Appl
33	85.5	10.4	1231	2	US-08-231-193A-48	Sequence 48, Appl
34	85.5	10.4	1231	2	US-08-486-273A-48	Sequence 48, Appl
35	85.5	10.4	1231	3	US-08-480-474-48	Sequence 48, Appl
36	85.5	10.4	1231	3	US-08-940-086A-48	Sequence 48, Appl
37	85.5	10.4	1236	2	US-08-231-193A-6	Sequence 6, Appl
38	85.5	10.4	1236	2	US-08-486-273A-6	Sequence 6, Appl
39	85.5	10.4	1236	3	US-08-480-474-6	Sequence 6, Appl
40	85.5	10.4	1236	3	US-08-940-086A-6	Sequence 6, Appl
41	85.5	10.4	1239	2	US-08-231-193A-52	Sequence 52, Appl
42	85.5	10.4	1239	2	US-08-486-273A-52	Sequence 52, Appl
43	85.5	10.4	1239	3	US-08-480-474-52	Sequence 52, Appl
44	85.5	10.4	1239	3	US-08-940-086A-52	Sequence 52, Appl
45	85.5	10.4	1244	2	US-08-231-193A-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-08-826-246-10
; Sequence 10, Application US/08826246
; Patent No. 6048709
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,246
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-078-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-826-246-10

Query Match 100.0%; Score 823; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 4e-75;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVVRQRLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120
DB 63 RSRRLVPRVVRQRLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNLTSEPSDYALDLSTFLOQHPPAAF 156
DB 121 APTVPSPVLEPFLNLTSEPSDYALDLSTFLOQHPPAAF 156

RESULT 2
US-08-944-495-10
; Sequence 10, Application US/08944495
; Patent No. 6087477
; GENERAL INFORMATION:
; APPLICANT: Faib, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONOS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,495
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-944-495-10

Query Match 100.0%; Score 823; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 4e-75;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVVRQRLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120
DB 61 RSRRLVPRVVRQRLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNLTSEPSDYALDLSTFLOQHPPAAF 156
DB 121 APTVPSPVLEPFLNLTSEPSDYALDLSTFLOQHPPAAF 156

RESULT 3
US-09-126-640-11
; Sequence 11, Application US/09126640A
; Patent No. 6099823
; GENERAL INFORMATION:
; APPLICANT: Faib, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE;
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/126,640A
; CURRENT FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 08/870,434
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 08/799,910
; EARLIER FILING DATE: 1997-02-13
; EARLIER APPLICATION NUMBER: 60/011,787
; EARLIER FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-126-640-11

Query Match 100.0%; Score 823; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 4e-75;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVVRQRLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120
DB 61 RSRRLVPRVVRQRLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNLTSEPSDYALDLSTFLOQHPPAAF 156
DB 121 APTVPSPVLEPFLNLTSEPSDYALDLSTFLOQHPPAAF 156

RESULT 4
US-09-035-648-24
; Sequence 24, Application US/09035648
; Patent No. 6100031
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL
; TITLE OF INVENTION: GROWTH AND PROLIFERATION
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/035,648
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/818,829
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 739 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-035-648-24

Query Match 11.6%; Score 95.5; DB 3; Length 739;

Best Local Similarity 27.0%; Pred. No. 0.089;

Matches 47; Conservative 14; Mismatches 60; Indels 53; Gaps 7;

QY 16 ATTPAPSTIPG-----RRSGPEITFTDPLPEPAAAPAGRPS-----ASRGHR---- 59
DB 264 APPYSAVTPPDATSRGVSSDIAGPAQPPWPQAPWSPAFYDSSERIASRDERISVP 323
QY 60 -KRS-----RVLYPRVVRQLPVEENPAKRLLFLL----- 91
DB 324 AKRTGILQAKRRSTTKPMFTKEPKVSPNP--ELISLLQNSEGRKCTGAGDGSPEEDY 381
QY 92 -----TIVFCQILMAEGVPAPLPEDAPNAAASLAP-TPVSPVLEPFLNTSEP 138
DB 382 LSLGAEACNFMSSSAKQKTPPVAPKPAVKSSSQVTPVSPVNSGVVAPTQP 435

RESULT 5

US-09-356-952-3

Sequence 3, Application US/09356952

Patent No. 6117663

GENERAL INFORMATION:

APPLICANT: Portack-Sjodin, Ann

APPLICANT: Margarit, S. M.

APPLICANT: Bor-Sogli, Dafna

APPLICANT: Cole, Philip

APPLICANT: Kuriyan, John

TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE

TITLE OF INVENTION: THEROF

FILE REFERENCE: 600-1-228N

CURRENT APPLICATION NUMBER: US/09/356,952

CURRENT FILING DATE: 1999-07-19

EARLIER APPLICATION NUMBER: 60/093,631

EARLIER FILING DATE: 1998-07-21

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 3

LENGTH: 1596

TYPE: PPT

ORGANISM: Drosophila melanogaster

US-09-356-952-3

Query Match

11.6%; Score 95.5; DB 3; Length 1596;

Best Local Similarity 28.3%; Pred. No. 0.22;

Matches 52; Conservative 19; Mismatches 68; Indels 45; Gaps 13;

QY 4 SRSCHPTWTILOAPTPAPSTIPGPRRG--SGPEITFTDPLPEPAAAPAGRPSASGHKKR 61
DB 1354 TESCADMAKQKQAP-DAP-TLP-PROGELSP-----PIPRNLNISTGISLRSUHGKS 1404
QY 62 SRRV-----LYPR-----VVRQLPVE-----ENPAKRLLFLLLTIVFCQILMAEGV 105
DB 1405 KEFVGNSLLPNTSSIMIRNSAIEKRAAATSQPNOAAAGPISTTLVTVSQAVATDEPL 1464
QY 106 PAPLPEDAPNAAASLAP--PVSPVLEP--NLTSPE-----SDYALDLSTLQO-----H 152
DB 1455 PLPISP-----AASSSTTSPPLTAPWSPNIPSHPVSTSSSYAHLRMHQOQQOQTH 1519
QY 153 PAAF 156
DB 1520 PAIY 1523

RESULT 6

US-08-574-959A-9

Sequence 9, Application US/08574959A

Patent No. 5962224

GENERAL INFORMATION:

APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi

APPLICANT: and Jack L. Strominger

TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES

TITLE OF INVENTION: AND USES THEREFOR

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESS: LAHIVE & COCKFIELD

STREET: 60 State Street, Suite 510

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA: US/08/574,959A

FILING DATE: 19-DEC-95

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: DFN-008

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 905 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-574-959A-9

Query Match

11.4%; Score 93.5; DB 2; Length 905;

Best Local Similarity 23.1%; Pred. No. 0.18;

Matches 45; Conservative 17; Mismatches 62; Indels 71; Gaps 8;

QY 8 HPTMTILO-----APTAPSTIPGPRRGSGPE-----ITFTDPLPEPAAAPAGRPSAS 55
DB 400 HPRVPPLOPMGFTCTPAPVLLRPHRPSGPHRSILRAPCOWAPCQOAPCPGAPMPS 459
QY 56 RG-----HRKRSRR--VLYP----- 68

DB 460 AGPVSEPTSTANLLGSLRSPVCPPLLLPGPENHRAGSNEDPILAPSGTPTPTIPP 519
QY 69 -----RVVRQLPVEENPAKRLLFLLLTIVFCQILMAEGVPAPLPEDAPNAAASLAP 123
DB 520 ETFGGRVPRPAFVHYDKKEASDVETSLSDSDSVVIVPEGLP-PLPP---PPPSGATPP 575

QY 124 PVSVPLEPNTSEP 138

Db 576 PIAPGPP---TASP 587

RESULT 7

US-08-574-959A-7

; Sequence 7, Application US/08574959A

; Patent No. 5962224

; GENERAL INFORMATION:

; APPLICANT: Jaekyoon Shin, Insil Jounq, Ratna K. Vadlamudi

; APPLICANT: and Jack L. Strominger

; TITLE OF INVENTION: p62 STROMINGER, RELATED POLYPEPTIDES

; TITLE OF INVENTION: AND USES THEREFOR

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 60 State Street, Suite 510

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/574.959A

; FILING DATE: 19-DEC-95

; ATTORNEY/AGENT INFORMATION:

; NAME: Mandragouras, Amy E.

; REGISTRATION NUMBER: 36,207

; REFERENCE/DOCKET NUMBER: DFN-008

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1135 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-574-959A-7

Query Match

Best Local Similarity 11.4%; Score 93.5; DB 2; Length 1135;

Matches 45; Conservative 17; Mismatches 62; Indels 71; Gaps 8;

QY 8 HPTMTILQ-----ATPAPSTIPGRGSGPE-----IFTFDPLPEAAAPAGRPAS 55

Db 630 HPRVPLQPMGTCTTPAPVILLRPHRSGPHRSILRAPCPQWAPCPQACPSAGPWS 689

QY 56 RG-----HRKRSRR--VLYP----- 68

Db 690 AGVPSEPTWTANLLGLLSRPSVCPRLPLPGPENHRAGSNEDDILAPSGTPTTPD 749

QY 69 -----KVRROLVPEPNAPKRLLELLTIVFCQILMAEEGVAPILPPEDAPNAASLAPT 123

Db 750 ETFGRVPRPAPVHYDKAESDVEISLESDSDSVIVVEGLP-PLPP---PPPSGATPP 805

QY 124 PVSVPLEPNTSEP 138

Db 806 PIAPGPP---TASP 817

RESULT 8

US-08-080-255-7

; Sequence 7, Application US/08080255

; Patent No. 548970

; GENERAL INFORMATION:

; APPLICANT: Rowley, Janet D.

; APPLICANT: Diaz, Manuel O.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND

; TITLE OF INVENTION: TRANSLOCATIONS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P. O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: USA

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/080.255

; FILING DATE: 19930617

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Parker, David L.

; REGISTRATION NUMBER: 32,165

; REFERENCE/DOCKET NUMBER: ARCD:072/PAR

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 320-7200

; TELEFAX: (512) 474-7577

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1400 amino acids

; TYPE: AMINO ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-080-255-7

Query Match

Best Local Similarity 11.2%; Score 92; DB 1; Length 1400;

Matches 31; Conservative 17; Mismatches 40; Indels 52; Gaps 5;

QY 15 QAPTAPSTIPGRGSGPEIFTDPLPE-----PAAAPAGRPASAGHKKKS 62

Db 465 QKPTPSAREDPAPKKSS-----EPPRPKPVEEKSEGNVSAPGPESKQATTPASRKSS 518

QY 63 RRVLYPRV-----RRQLPVEBPNAKRLLELLTIVFCQILMAEEGVAPILPP 111

Db 519 KQVQPALVIPQPTTCTPPRKEVKTTPSEKK-----KQPPPP 558

QY 112 EDAPNAA---SLAPTPVSPV 128

Db 559 ESGPEQSKQKVAAPRPSIPV 578

RESULT 9

US-08-465-713-7

; Sequence 7, Application US/08465713

; Patent No. 6121419

; GENERAL INFORMATION:

; APPLICANT: Rowley, Janet D.

; APPLICANT: Diaz, Manuel O.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND

; TITLE OF INVENTION: TRANSLOCATIONS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P. O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: USA

LENGTH: 214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-217-327-4

Query Match
Best Local Similarity 27.5%; DB 1; Length 214;
Matches 38; Conservative 8; Mismatches 59; Indels 33; Gaps 5;

QY 9 PTMTLQATPIAPSTIPGRRGSGEITFDPI-----LPEAAAPAGRPSASRCHKRKRS 62
DB 54 PTATPPPVSTPTPTTPPPVSTPTTPPPPPATPTPPASPPPTATPTPPASPPPTATP 113
QY 63 RRVLYPRVVRQLPVEENPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASLAP 122
DB 114 PPASPPATP---PPATPPA-----TPPPATPPAPLASPPAT 149
QY 123 TP-VSPVLEPFLTSEPS 139
DB 150 VPAISPQVTP--LTSPPA 165

RESULT 12

US-08-390-000A-7
Sequence 7, Application US/08390000A
Patent No. 5985583

GENERAL INFORMATION:
APPLICANT: Sealion, Stuart C.
TITLE OF INVENTION: Cloning and Expression of
TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,000A
FILING DATE: 17-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6923-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-390-000A-7

Query Match
Best Local Similarity 24.2%; DB 2; Length 468;
Matches 50; Conservative 19; Mismatches 44; Indels 94; Gaps 12;

QY 17 PTPASTIPGRRGSGEITFDPLPEAAAPAGRPSAS--RGHRKRSRVLYPRVVRQ 74
DB 264 PPSPPSVPTAPAPPPCP-----PRPAAAAATAPLANGRAGRPSRLV-----ALREQ 311

QY 75 -----LPVEENPAKRL-----LFLLLT-----LVFCO- 97
DB 312 KALKTLGIINGVFTLQNLFFFLANVVKAFHRELVPDLFVFFNWLGYANSAPNPILYCKRS 371
QY 98 --ILMAEGV-----PAPLPEDAPNAAS-----LA 121
DB 372 PDFRKAFOCLCCARRAARRRHATHGDRPRASGCLARPGI-PP--SIVGAASDDDDDUUVG 428
QY 122 PTPVSPVLEPF-----NLTSEPSDYALD 144
DB 429 ATPPARILLEHWACNGCGAAADSSSLD 455

RESULT 13

US-08-444-734A-4
Sequence 4, Application US/08444734A
Patent No. 5610282

GENERAL INFORMATION:
APPLICANT: Sibley, David R.
APPLICANT: Monsma, Frederick J.
APPLICANT: Mahan, Lawrence C.
APPLICANT: McVittie, Loris D.
TITLE OF INVENTION: cDNA encoding the rat D1 dopamine
TITLE OF INVENTION: receptor linked to adenylyl cyclase activation and
TITLE OF INVENTION: expression of the receptor protein in plasmid-transfected
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson and Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,734A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/029,917
FILING DATE: 03-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/548,714
FILING DATE: 06-JUL-1990

ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: NIH065.001FW1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-444-734A-4

Query Match
Best Local Similarity 10.7%; DB 1; Length 477;
Matches 50; Conservative 19; Mismatches 44; Indels 94; Gaps 12;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 18:47:31 : Search time 54.97 Seconds
(without alignments)
97.039 Million cell updates/sec

Title: US-08-799-910-10
Perfect score: 823
Sequence: 1 MCHSRCHPTMTILOAFTP.....EPDYALDLSTLQHPAAF 156

Scoring table: BLCSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36.*
1: /cgn2_2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /cgn2_2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /cgn2_2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /cgn2_2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /cgn2_2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /cgn2_2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /cgn2_2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /cgn2_2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /cgn2_2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /cgn2_2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /cgn2_2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /cgn2_2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /cgn2_2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /cgn2_2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /cgn2_2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /cgn2_2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /cgn2_2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT.*

prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result, being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	823	100.0	156	21	Y45017 Protein encoded by
2	816	99.1	156	18	W36006 Human fchd605 gene
3	107	13.0	206	18	W14574 Streptococcus pneu
4	104	12.6	129	18	W14573 Streptococcus pneu
5	101.5	12.3	190	18	W14569 Streptococcus pneu
6	101	12.3	180	18	W14562 Streptococcus pneu
7	101	12.3	183	18	W14570 Streptococcus pneu
8	100	12.2	783	19	W37151 Mouse neural Mena+
9	100	12.2	787	19	W37152 Mouse neural Mena+
10	100	12.2	802	19	W37153 Mouse neural Mena+
11	98.5	12.0	168	21	Y68732 Amino acid sequenc
12	96.5	11.7	179	21	Y68731 Amino acid sequenc

13	95.5	11.6	739	18	W25790
14	95.5	11.6	1596	21	Y68821
15	94.5	11.5	520	16	R78185
16	94	11.4	1290	20	Y06809
17	94	11.4	1290	21	Y32216
18	94	11.4	1291	20	Y01687
19	93.5	11.4	185	18	W14566
20	93.5	11.4	905	18	W31186
21	93.5	11.4	1135	18	W31185
22	92.5	11.2	572	18	W31855
23	92.5	11.2	763	18	W31852
24	92	11.2	166	18	W14568
25	92	11.2	167	18	W14575
26	92	11.2	1400	15	R44514
27	92	11.2	1400	17	R92705
28	92	11.2	3910	14	R38470
29	92	11.2	3910	16	R66462
30	91	11.1	1012	20	Y17406
31	90.5	11.0	214	17	R86913
32	90	10.9	171	20	Y29034
33	90	10.9	171	20	Y29074
34	90	10.9	254	20	Y29061
35	89.5	10.9	666	19	W72911
36	89.5	10.9	666	20	Y21928
37	89	10.8	788	21	Y54466
38	89	10.8	807	21	Y54467
39	89	10.8	1291	20	Y01689
40	89	10.8	1291	20	Y06810
41	89	10.8	3969	15	R52971
42	88.5	10.8	187	18	W14579
43	88.5	10.8	188	18	W14580
44	88.5	10.8	688	21	Y57898
45	88.5	10.8	711	19	W80602

ALIGNMENTS

RESULT 1

ID Y45017 standard; Protein; 156 AA.

XX AC Y45017;

DT 31-MAY-2000 (first entry)

DE Protein encoded by fchd605 gene.

XX fchd605 gene; human; cardiovascular disease; oncogenic disorder;

KW diabetic retinopathy; fibroproliferative disorder; atherosclerosis;

KW TGF-beta signalling pathway; TGF; Transforming growth factor;

KW pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;

KW vascularisation; cytostatic; antidiabetic; ophthalmological.

OS Homo sapiens.

XX WO200006206-A1.

PD 10-FEB-2000.

XX PF 30-JUL-1999; 99WO-US17394.

XX PR 30-JUL-1998; 98US-0126640.

XX PA (MILL-) MILLENNium PHARM INC.

XX PI Falb DA;

XX WPI; 2000-205414/18.

XX N-PSDB; 250711.

PT Identifying substances for ameliorating symptoms of fibroproliferative diseases or oncogenic related disorders -

XX

Example; Fig 5; 214pp; English.

XX

The patent discloses methods for the treatment and diagnosis of cardiovascular diseases by novel human genes which are differentially expressed in different cardiovascular disease states. Compositions which can modify TGF-beta signalling pathway are identified by screening. These are used therapeutically to treat fibroproliferative and oncogenic disorders, especially TGF (transforming growth factor)-beta related disorders, including diabetic retinopathy, arteriosclerosis, pancreatic cancer, angiogenesis, inflammation, fibrosis, tumour growth and vascularisation. The present sequence is the protein product of fchd605 gene which is up-regulated in monocytes treated with oxidised LDL (low density lipoprotein) can be used to design cardiovascular disease treatment strategies. Depending on whether the up-regulation has a pathogenic or protective effect treatment methods can be designed to increase or decrease the activity of the protein product of the gene.

XX Sequence 156 AA;

Query Match 100.0%; Score 823; DB 21; Length 156;
 Best Local Similarity 100.0%; Pred. No. 7.4e-69;
 Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPASRGRK 60
 Db 1 mchsrchptmtlilqaptapstipgprgsgpeitfdplpepaaapagrpasrgrk 60
 OY 61 RSRRLVYPRVVRQLPVEEPNPAKRLLFLITIVFCQILMAEGYVAPILPPEDAPNAASL 120
 Db 61 rsrriyprvrrqlpveepnpakrllflitivfcqilmaegypapilppedapnaasl 120
 OY 121 APTPSVPLEPNLTSEPSDYALDLSTFLOQHPAAF 156
 Db 121 aptpsvplepnltspsdyaldstflqhpaaaf 156

RESULT 2

W36006
 ID W36006 standard; Protein; 156 AA.

XX W36006;

XX 03-MAR-1998 (first entry)

XX Human Fchd605 gene product.

XX Fchd605 gene; differential expression; monocyte; human;
 KW foam cell; cardiovascular disease; arteriosclerosis; ischaemia;
 KW reperfusion; hypertension; restenosis; arterial inflammation;
 KW therapy; diagnosis; drug screening; marker.

XX Homo sapiens.

XX WO9730065-A1.

XX 21-AUG-1997.

XX 14-FEB-1997; 97WO-US02291.

XX 13-FEB-1997; 97US-0799910.

XX 16-FEB-1996; 96US-0011787.

XX (MILL-) MILLENNIUM PHARM INC.

XX Fatb DA;

XX WPI; 1997-424966/39.

XX N-PSDB; T94471.

XX New genes differentially expressed in cardiovascular disease - used
 for diagnosis, drug screening and treatment of cardiovascular

PT disease, e.g. atherosclerosis, restenosis, hypertension, etc

XX Example 6; Fig 5; 163pp; English.

XX This protein is encoded by the novel human fchd605 gene (see
 CC T94471) that is up-regulated in monocytes treated with oxidised
 CC low density lipoproteins that simulate the conditions under which
 CC foam cells develop during atherogenesis. The protein has sequence
 CC similarity to the mouse gly96 gene and to EST T49532. Novel
 CC fchd531, fchd540, fchd545, fchd602 and fchd605 genes (see T94467-71)
 CC provide a fingerprint for the study of cardiovascular diseases,
 CC including atherosclerosis, ischaemia/reperfusion, hypertension,
 CC restenosis and arterial inflammation. Methods are provided for the
 CC diagnosis, monitoring in clinical trials, screening for
 CC therapeutically effective compounds, and treatment of
 CC cardiovascular diseases based on discoveries regarding the
 CC expression patterns of these novel genes.

XX Sequence 156 AA;

Query Match 99.1%; Score 816; DB 18; Length 156;
 Best Local Similarity 99.4%; Pred. No. 3.3e-68;
 Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPASRGRK 60
 Db 1 mchsrchptmtlilqaptapstipgprgsgpeitfdplpepaaapagrpasrgrk 60
 OY 61 RSRRLVYPRVVRQLPVEEPNPAKRLLFLITIVFCQILMAEGYVAPILPPEDAPNAASL 120
 Db 61 rsrriyprvrrqlpveepnpakrllflitivfcqilmaegypapilppedapnaasl 120
 OY 121 APTPSVPLEPNLTSEPSDYALDLSTFLOQHPAAF 156
 Db 121 aptpsvplepnltspsdyaldstflqhpaaaf 156

RESULT 3

W14574
 ID W14574 standard; Protein; 206 AA.

XX W14574;

XX 28-OCT-1997 (first entry)

XX Streptococcus pneumoniae PspA central region.

XX PspA; pneumococcal surface protein; vaccine; otitis media;
 KW meningitis; bacteraemia; pneumonia.

XX Streptococcus pneumoniae strain Db15.

XX Key location/Qualifiers

FT Misc-difference 50
 FT /note= "unidentified amino acid"

XX WO9709994-A1.

XX 20-MAR-1997.

XX 16-SEP-1996; 96WO-US14819.

XX 15-SEP-1995; 95US-0529055.

XX (UARR-) UAB RES FOUND.

XX Briles DE, Brooks-walter A, Crain MJ, Hollingshead S;
 PI McDaniel LS, Swiatlo E, Tarl R, Yother J;

XX WPI; 1997-202002/18.

XX Streptococcus pneumoniae surface protein pspC and truncated pspA -
 PT

xx
ps
xx

Example 6; Fig 13; 296pp: English.

CC This sequence shows the central portion, including the C-terminus
CC of the alpha-helix region and some of the proline-rich region, of
CC pneumococcal surface protein A (PspA) of *Streptococcus pneumoniae*
CC strain E6796. Comparison of the N-terminal and central regions
CC (W1453-57 and W1562-91) of PspA polypeptides from different
CC pneumococcal strains can be used to divide the strains into several
CC families based on sequence homologies. PspA polypeptides, or
CC fragments of them, can be used in vaccines to protect animals
CC against *S. pneumoniae* infection and hence for the prevention of
CC diseases such as otitis media, meningitis, bacteraemia and pneumonia.
CC The sequence of the 3' half of the PspA alpha-helical region and the
CC immediate 5' tip of the coding sequence are likely to be the critical
CC sequences for predicting PspA cross-reactions and vaccine
CC composition.

xx Sequence 129 AA:
xs

```
Query Match      12.6%; Score 104; DB 18; Length 129;  
Best Local Similarity 26.5%; Pred. No. 0.012;  
Matches 30; Conservative 8; Mismatches 25; Indels 50; Gaps 4;  
  
QY 16 APTAPSTIFGPRGSGPEFTFUPLEPAAAAAGRVSASGHRKSRVLYPVVRKQL 75  
   ||||| : | : | ||| ||| :  
Db 64 apapapapapkpapapk-----papapapapkpa----- 96  
  
QY 76 PVEEPNTAKRLLELLLTITVCQILMAEEGVAPALP-PEDAPNAASSLAPTVPSP 127  
   | : || : ||| ||| : ||| : ||| :  
Db 97 papkapa-----papkpckbackpapakpetp 127
```

RESULT	5	
W14569		
ID	W14569 standard; Protein; 190 AA.	
XX		
AC	W14569;	
XX		
DT	28-OCT-1997 (first entry)	
XX		
DE	Streptococcus pneumoniae PspA central region.	
XX		
XX	PspA; pneumococcal surface protein; vaccine; otitis media;	
KW	meningitis; bacteraemia; pneumonia.	
KW		
XX		
OS	Streptococcus pneumoniae strain Bg8743.	
XX		
XX		
PH	Key	Location/Qualifiers
FT	Misc-difference 21	
FT	/note= "unidentified amino acid"	
FT	Misc-difference 24	
FT	/note= "unidentified amino acid"	
FT	Misc-difference 95	
FT	/note= "unidentified amino acid"	
FT	Misc-difference 97	
FT	/note= "unidentified amino acid"	
FT	Misc-difference 186	
FT	/note= "unidentified amino acid"	
XX		
XX	WO9709994-A1.	
PN		
XX		
PD	20-MAR-1997.	
PD		
XX		
PF	16-SEP-1996; 96WO-US14819.	
XX		
PR	15-SEP-1995; 95US-0529055.	
XX		
PA	(UABR-) UAB RES FOUND.	
XX		
PI	Briles DE, Brooks-Walter A, Crain MJ, Hollingshead S;	
PI	McDaniel LS, Swiatlo E, Tart R, Yother J;	

CC This protein comprises novel murine neural Meuα+. Its amino acid
CC sequence was deduced from a cDNA clone (see V02598) obtained from
CC

Example 4: Page 60-63: 77pp: English.

Example 4: Page 60-63: 77pp: English.

CC ligand dependent manner to induce gene expression can be identified
 CC using the method of the invention. The specification describes a
 CC method for identifying compounds that modulate a nuclear receptor.
 CC The method comprises incubating a polypeptide containing a
 CC ligand-binding region of nuclear receptor with test compound in
 CC presence of second polypeptide containing the nuclear receptor-binding
 CC part of a HMG protein, and detecting any alteration in binding between
 CC the two polypeptides. The method is useful for identifying potential
 CC antiproliferative, anti-inflammatory, antipsoriasis and anticancer
 CC agents.
 XX
 SQ Sequence 168 AA;

Query Match 12.0%; Score 98.5; DB 21; Length 168;
 Best Local Similarity 27.7%; Pred. No. 0.05;
 Matches 39; Conservative 9; Mismatches 52; Indels 41; Gaps 4;
 QY 17 PTPAPSTIPGRRSGPEITFDPLPEPAAPAGRSASGHRKRRLVPRVRL- 75
 DB 33 PKKPSVPTPKR-----prgpkyskknwrrkrasrrsprts 74
 QY 76 -PVEEPNPAKRLFLLLTVFCQILMAEEGVAPLPEDAPNAASLAPTPVSPVLEPFNL 134
 DB 75 dpcvpapahwrssflglgsf-----apipp-----ppplpqahhhrl 113
 QY 135 TSEPSDYALDSTFLQHPAA 155
 DB 114 wpppsstcaittlhtspaa 134

RESULT 12
 Y68731
 ID Y68731 standard; Protein; 179 AA.
 AC Y68731;
 DT 05-MAY-2000 (first entry)
 DE Amino acid sequence of high mobility group (HMG)-R protein.
 XX
 KW Retinoic acid receptor; RAR; high mobility group protein;
 KW HMG protein; nuclear receptor; antiproliferative; anti-inflammatory;
 KW antipsoriasis; anticancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200001717-A2.
 PD 13-JAN-2000.
 XX
 PF 28-JUN-1999; 99WO-US14801.
 XX
 PR 01-JUL-1998; 98US-0108298.
 XX
 PA (ALLR) ALLERGAN SALES INC.
 XX
 PI Nagpal S, Chandratatna RA, Chosn C;
 XX
 DR WPI: 2000-170999/15.
 XX
 PT Identifying modulators of nuclear receptors, useful e.g. as anticancer
 PT and anti-inflammatory agents, according to their effect on binding
 PT between the receptor and high-mobility group protein.
 XX
 PS Claim 24; Fig 1B; 43pp; English.
 XX
 CC The present sequence represents a high mobility group (HMG)-R protein.
 CC HMG sequences that interact with retinoic acid receptors (RARs) in a
 CC ligand dependent manner to induce gene expression can be identified
 CC using the method of the invention. The specification describes a
 CC method for identifying compounds that modulate a nuclear receptor.
 CC The method comprises incubating a polypeptide containing a

CC ligand-binding region of nuclear receptor with test compound in
 CC presence of second polypeptide containing the nuclear receptor-binding
 CC part of a HMG protein, and detecting any alteration in binding between
 CC the two polypeptides. The method is useful for identifying potential
 CC antiproliferative, anti-inflammatory, antipsoriasis and anticancer
 CC agents.
 XX
 SQ Sequence 179 AA;

Query Match 11.7%; Score 96.5; DB 21; Length 179;
 Best Local Similarity 27.4%; Pred. No. 0.083;
 Matches 40; Conservative 9; Mismatches 56; Indels 41; Gaps 4;
 QY 12 TILQAPTPADSTIPGRRSGPEITFDPLPEPAAPAGRSASGHRKRRLVPRVV 71
 DB 39 talvsgskpsevtptkr-----prgpkyskknwrrkrasrrsp 80
 QY 72 RRQL--PVEEPNPAKRLFLLLTVFCQILMAEEGVAPLPEDAPNAASLAPTPVSPVL 129
 DB 81 rrrssdpcvppahwrssflglgsf-----apipp-----ppplpqah 119
 QY 130 EPNLTSSEPSDYALDSTFLQHPAA 155
 DB 120 hhhrlwpppsstcaittlhtspaa 145

RESULT 13
 W25790
 ID W25790 standard; Protein; 739 AA.
 AC W25790;
 DT 11-MAR-1998 (first entry)
 DE Gene 036 product differentially expressed in colon tumour cells.
 XX
 KW Colon tumour; colon cancer; differential expression; gene 036;
 KW human; diagnosis; gene therapy; tumour suppressor.
 XX
 OS Homo sapiens.
 XX
 PN WO9733551-A2.
 PD 18-SEP-1997.
 XX
 PF 14-MAR-1997; 97WO-US04191.
 XX
 PR 15-MAR-1996; 96US-0013438.
 XX
 PA (MILL-) MILLENNIUM PHARM.
 PI Shyjan AW;
 XX
 DR WPI: 1997-470615/43.
 DR N-PSDB; T91708.
 XX
 PT Gene 036 with reduced or amplified expression in tumour cells -
 PT used, optionally with genes 097, 030 and 056 or their protein
 PT products, for diagnosis and treatment of colonic cancer
 XX
 PS Claim 38; Fig 2; 117pp; English.
 XX
 CC This protein is encoded by gene 036 (see T91708), which was
 CC identified by differential display analysis as being expressed at
 CC a higher level in normal colon tissues than in cancerous colon
 CC tissues. Gene 036 is a candidate tumour suppressor gene. A
 CC correlation was found between an increase in the expression level
 CC of gene 036 and a decrease in a colon cell's tumour potential.
 CC Hence, methods that increase the level of expression of gene 036
 CC may inhibit or slow the progression to tumours and cancer, e.g.
 CC colon cancer. The 036 protein is used in claimed methods for
 CC treating a patient suffering from a disorder associated with

CC Insufficient expression of gene 036 protein, and for identifying
 CC compounds that modulate 036 protein activity. Such compounds are
 CC useful in the diagnosis, prevention and treatment of tumours and
 CC cancers.

SQ Sequence 739 AA;

Query Match 11.6%; Score 95.5; DB 18; Length 739;
 Best Local Similarity 27.0%; Pred. No. 0.5;
 Matches 47; Conservative 14; Mismatches 60; Indels 53; Gaps 7;
 QY 16 APTAPSTIPG-----RRSGPEITFDPLPEPAAAPAGRPS-----ASRGHR----- 59
 DB 264 appysavtpdpdafsgvsggiagpqpqpwpqapwsqpfydsrseriasrderisvp 323
 QY 60 -RRS-----RVLYPRVVRQLPVEPNPAKRLFLLL----- 91
 DB 324 akrtgilqakrstlckpmftfkepkvsnp--ellslqnsegkrgtgaggdsgpeedy 381
 QY 92 -----TIVFCQILMAERGVPAIPPEPAPNAASLAP--TPVSPVLEPFWMTSEP 138
 DB 382 lslgaecnfmdjssakqtkppvpakpvaksssqvtpvpsvwpvgvaptqp 435

RESULT 14

Y68821
 ID Y68821 standard; protein: 1596 AA.

XX AC Y68821;

DT 16-MAY-2000 (first entry)

XX Amino acid sequence of a Drosophila Son of sevenless (Sos) protein.

XX Ras; Son of sevenless; Sos; crystal; Ras-Sos complex; cancer;

XX protein coordinate data.

XX Drosophila melanogaster.

XX WO200005258-A1.

XX 03-FEB-2000.

XX 20-JUL-1999; 99WO-US16348.

XX 21-JUL-1998; 98US-0119794.

XX (UYRQ) UNIV ROCKEFELLER.

XX Horiack-Sjodin A, Margarit SM, Bar-Sagi D, Cole P, Kuriyan J;

XX WP1; 2000-182647/16.

XX Novel crystals comprising a Ras-Son of sevenless complex, useful for
 PT screening drugs useful in cancer treatment

XX Disclosure; Page 169-175; 224pp; English.

XX The specification describes a crystal complex comprising at least a
 CC Sos contacting region of a Ras protein and at least a Son of sevenless
 CC (Sos) protein catalytic region fragment, that effectively diffracts
 CC X-rays. Ras and Sos form a tight complex. Sos does not impede the
 CC binding sites for the nucleotide base and the ribose of GTP or GDP
 CC and thus the Ras-Sos complex maintains a structure that permits
 CC nucleotide release and rebinding. The crystals are used for the
 CC determination of the atomic coordinates of the complex to a resolution
 CC of more than 5.0 Angstrom. The crystals, or a dataset comprising the
 CC three-dimensional coordinates obtained from the crystals, is useful
 CC for identifying an agent that stabilizes the Ras-Sos complex. The
 CC crystals are also useful for identifying agents that inhibit the
 CC formation of Ras-Sos complex. Ras and Sos fragments are useful for
 CC growing a crystal of a protein-ligand complex. Agents that stabilize

CC or inhibit the formation of Ras-Sos complex are useful in the
 CC treatment of cancer. The present sequence represents a Drosophila
 CC Sos protein.

SQ Sequence 1596 AA;

Query Match 11.6%; Score 95.5; DB 21; Length 1596;
 Best Local Similarity 28.3%; Pred. No. 1.2;
 Matches 52; Conservative 19; Mismatches 68; Indels 45; Gaps 13;
 QY 4 SRSHPTMTILQAPTAPSTIPGPRRG--SGPEITFDPLPEPAAAPAGRPSASRGHKKR 61
 DB 1354 tescadmakqkap-dap-tlp-prdgelspp-----piprlnhstgisylrqshoks 1404
 QY 62 SRKV-----LYPR-----VVRQLPVE-----EPNPAKRLFLLLTIVFCQILMAE 105
 DB 1405 kefvgnssllilpnrtsslmfrnsalekrraaatsqpnqaaqpisttlvlvsqavatdepl 1464
 QY 106 PAPLPPEPAPNAASLAP--PVSPVLEPF--NLTSEP-----SDYALDLSTFLQO----H 152
 DB 1465 pipisp-----aassttsplitpamspmspnipspvestsssyahqlrmrqggqgth 1519
 QY 153 PAAF 156
 DB 1520 paiv 1523

RESULT 15

R78185
 ID R78185 standard; Protein: 520 AA.

XX AC R78185;

DT 09-FEB-1996 (first entry)

XX Protein sequence of PEA3-beta -an ETS transcription factor.

XX Transcription factor; probe; reverse transcription; PCR; primer;

XX expression vector; E.coli; COS cell; ras; cancer cell multiplication;

XX polyoma virus; transformation.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 84 /note= "encoded by CTC"

FT Misc-difference 126 /note= "encoded by AAC"

FT Misc-difference 500 /note= "encoded by ACC"

XX JP07145197-A.

XX 06-JUN-1995.

XX 25-NOV-1993; 93JP-0295393.

XX 25-NOV-1993; 93JP-0295393.

XX (EISA) EISAI CO LTD.

XX (HIRA/) HIRANO T.

XX WPI; 1995-237197/31.

XX N-PSDB; Q91769.

XX ETS transcription factor activated by ras - may be used in the study
 PT of cancer cell proliferation and the proliferation of the polyoma
 PT virus

XX Claim 1; Page 7-9; 9pp; Japanese.

XX The amino acid sequence of the novel ETS transcription factor family

member - pEA3-beta. The gene was isolated from a HepG2 cell line cDNA library. The probe for the screening was prepared by reverse transcription on Hep2 mRNA followed by PCR using primers G91770-1, to produce a probe of 170-200 bp. The screening isolated the full length sequence of the transcription factor. The gene was inserted into the expression vectors pBluescript KS and pCDV1 for expression of the protein in E.coli and COS7 cells, respectively. The ETS transcription factor has specificity for and is activated by ras. It is useful as a reagent in studies for the elucidation of the mechanism of cancer cell multiplication or polyoma virus transformation of cells.

sequence 520 AA;

Query Match	11.5%	Score 94.5;	DB 16;	Length 520;
Best Local Similarity	22.4%	Pred. No. 0.42;		
Matches	38;	Conservative 13;	Mismatches 62;	Indels 57; Gaps 7;
OY	3	HRSRCHPTMTILQAPTPAPSTIPGPRRGSGPEIFTDPLPEPAAAGAPGPSASRGIIHKRS	62	
Ddb	162	hapaaqpvvgv--gkapahsisipee---gpqqgtf-avprpbhqplqm-----kmm	207	
OY	63	RVLYPRVVRROLPVVEFPNPAKRLFLLLTVFCQILMAEGVP-----	106	
Ddb	208	penqlpysqrfrqlsep-----chfpqpqvpgdnrpsyrqmqsepi	252	
OY	107	--A'LPPED-----APNAASLAPTVPSPVLFPFNLITSDPSIYALD	144	
Ddb	253	paaippppqfgkqeyhdpl'yehavvpmpgppealgfqspmgikqepmdydcvd	302	

Search completed: January 30, 2001, 22:05:37
Job time: 11886 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 18:47:25 ; Search time 1450.12 Seconds
(without alignments)
1246.746 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468

Perfect score: 258

Sequence: 1 GTCCGGCCGACGTGCCACT.....AGCAACACCCGCCGCTTC 258

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
 118: em_estpl7:*
 119: em_estpl8:*
 120: em_estro14:*
 121: em_estro15:*
 122: em_estro16:*
 123: em_estro17:*
 124: em_estro18:*
 125: em_estro19:*
 126: gb_est58:*
 127: gb_est59:*
 128: gb_est60:*
 129: gb_est61:*
 130: gb_est62:*
 131: gb_est63:*
 132: gb_est64:*
 133: gb_est65:*
 134: gb_est66:*
 135: gb_est75:*
 136: gb_est76:*
 137: gb_est77:*
 138: gb_est78:*
 139: gb_est79:*
 140: gb_est80:*
 141: gb_est81:*
 142: gb_est82:*
 143: gb_est83:*
 144: gb_est84:*
 145: gb_est85:*
 146: gb_est86:*
 147: gb_est87:*
 148: gb_gss1:*
 149: gb_gss2:*
 150: gb_gss3:*
 151: gb_gss4:*
 152: em_gss1:*
 153: em_gss2:*
 154: em_gss3:*
 155: em_gss4:*
 156: gb_gss5:*
 157: gb_gss6:*
 158: gb_gss7:*
 159: gb_gss8:*
 160: gb_gss9:*
 161: gb_gss10:*
 162: gb_gss11:*
 163: gb_gss12:*
 164: gb_gss13:*
 165: gb_gss14:*
 166: gb_gss15:*
 167: gb_gss16:*
 168: gb_gss17:*
 169: gb_gss18:*
 170: gb_gss19:*
 171: gb_gss20:*
 172: gb_gss21:*
 173: gb_gss22:*
 174: gb_gss23:*
 175: gb_gss24:*
 176: em_gss5:*
 177: em_gss6:*
 178: em_gss7:*
 179: em_gss8:*
 180: em_gss9:*
 181: em_gss10:*
 182: em_gss11:*
 183: em_gss12:*
 184: em_gss13:*
 185: em_gss14:*
 186: em_gss15:*
 187: em_gss16:*
 188: em_gss17:*
 189: em_gss18:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	258	100.0	411	25	A1824906	A1824906 wb03e06.x
2	258	100.0	544	107	BE384949	BE384949 601276496
3	258	100.0	951	107	BE383865	BE383865 601273192
4	256.4	99.4	765	106	Bu275966	BE275966 601120727
5	254.4	98.6	762	110	H6621346	BE621346 601493736
6	249.4	96.7	916	137	H8882130	BE882130 601505227
7	245.4	95.1	603	106	BE273386	BE273386 601143493
8	244.6	94.8	463	7	AA410666	AA410666 zL30q09.r
9	242	93.8	585	97	AW960290	AW960290 EST372361
10	236.8	91.8	810	109	H5563592	BE563592 601334706
11	223	86.4	447	147	W52268	W52268 zc46a10.r1
12	217.2	84.2	513	11	AA743819	AA743819 OB01d04.s
13	205.4	79.6	212	142	H80073	H80073 yv80q11.r1
14	203	78.7	1070	137	BE878459	BE878459 601487866
15	197.8	76.7	504	2	AA143404	AA143404 zo66e09.r
16	195.6	75.8	502	90	AW465331	AW465331 B0230018B
17	195.6	75.8	523	135	BE751344	BE751344 203178 MA
18	195.6	75.8	547	135	BE750116	BE750116 201294 MA
19	195.6	75.8	557	139	BF040186	BF040186 BP250005A
20	195.6	75.8	571	139	BF045952	BF045952 BP250003B
21	195.6	75.8	594	37	AV617905	AV617905 AV617905
22	192	74.4	346	38	AV683186	AV683186 AV683186
23	189.2	73.3	545	139	BF043065	BF043065 BP2500026H
24	188.8	73.2	617	38	AV690660	AV690660 AV690660
25	181	70.2	184	143	N57203	N57203 yw91h09.r1
26	173.6	67.3	701	135	BE735733	BE735733 601304556
27	166	64.3	389	146	W39753	W39753 zc80b07.r1
28	163.8	63.5	174	137	BE939840	BE939840 RC6-UT001
29	160.8	62.3	597	37	AV594384	AV594384 AV594384
30	155	60.1	443	147	W47587	W47587 zc35b02.r1
31	155	60.1	806	139	BF025858	BF025858 601670012
32	154.4	59.8	628	96	AW916683	AW916683 EST347987
33	151.2	58.6	467	19	AI326484	AI326484 mq23c12.y
34	146.4	56.7	182	6	AA356940	AA356940 EST65572
35	145.8	56.5	150	89	AW367127	AW367127 MR0-HT015
36	144.8	56.1	487	139	BF023106	BF023106 ux04d04.y
37	143.2	55.5	471	109	BE554106	BE554106 ur38c03.y
38	139.4	54.0	1155	135	BE786915	BE786915 601477783
39	139.2	54.0	385	147	W77628	W77628 me68a09.r1
40	139.2	54.0	460	94	AW740625	AW740625 ur02h11.y
41	137	53.1	353	142	N45391	N45391 yw97d08.r1
42	136	52.7	425	2	AA122977	AA122977 mq23c12.r
43	127.6	49.5	443	147	W71619	W71619 me40d11.r1
44	126	48.8	459	26	AI876358	AI876358 uk74a05.y
45	125.4	48.6	476	12	AA833402	AA833402 ub58b04.r

ALIGNMENTS

RESULT 1
 A1824906/C

LOCUS A1824906 411 bp mRNA EST 16-DNC-1999
 DEFINITION wb03e06.x1 NC1_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2304610 3' similar to SW:IEH1_HUMAN P4695 RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 ; mRNA sequence.
 ACCESSION A1824906
 VERSION A1824906.1 GI:5445577
 KEYWORDS EST.

mRNA sequence.

ACCESSION BE383865
 VERSION BE383865.1 GI:9329230
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 951)

NH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

AUTHORS

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC/DCTD/DRP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LUCM276 row: 9 column: 24
 High quality sequence stop: 747.

FEATURES

source

1..951

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3614255"

/clone_lib="NIH_MGC_20"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

210 a 324 c 253 g 164 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 258; DB 107; Length 951;

Best Local Similarity 100.0%; Pred. No. 1.1e-55;

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGAACCCGAGCCAAAGGCTTCTTTCTGCTG 60

Db 210 GTCCGGCGCCAGCTGCCAGTCGAGAACCCGAGCCAAAGGCTTCTTTCTGCTG 269

Qy 61 CTCACCATCGTCTTCGCGCATCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCTT 120

Db 270 CTCACCATCGTCTTCGCGCATCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCTT 329

Qy 121 CCAGAGGAGCGCCCTAACGCGGCATCTGCTGGCGCCACCCCTGTGTCCCCCGTCTCGAG 180

Db 330 CCAGAGGAGCGCCCTAACGCGGCATCTGCTGGCGCCACCCCTGTGTCCCCCGTCTCGAG 389

Qy 181 CCCTTTAATCTGACTTGGAGCCCTCGAGTACGCTCTGGACCTTCAGACATTTCTCCAG 240

Db 390 CCCTTTAATCTGACTTGGAGCCCTCGAGTACGCTCTGGACCTTCAGACATTTCTCCAG 449

Qy 241 CAACACCGCGCGCTTC 258

Db 450 CAACACCGCGCGCTTC 467

RESULT

4

BE275966

LOCUS

DEFINITION 60110727F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2967534 5',
 mRNA sequence.

EST

765 bp

mRNA

13-JUL-2000

ACCESSION

BE275966

VERSION BE275966.1 GI:9150928

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 765)

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

AUTHORS

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC/DCTD/DRP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LUCM72 row: e column: 07
 High quality sequence start: 7
 High quality sequence stop: 699.

Location/Qualifiers

1..765

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2967534"

/clone_lib="NIH_MGC_20"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

151 a 284 c 202 g 128 t

BASE COUNT

ORIGIN

Query Match 99.4%; Score 256.4; DB 106; Length 765;

Best Local Similarity 99.6%; Pred. No. 2.8e-55;

Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGAACCCGAGCCAAAGGCTTCTTTCTGCTG 60

Db 257 GTCCGGCGCCAGCTGCCAGTCGAGAACCCGAGCCAAAGGCTTCTTTCTGCTG 316

Qy 61 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCGCCCTGCTT 120

Db 317 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCGCCCTGCTT 376

Qy 121 CCAGAGGAGCGCCCTAACGCGGCATCTGCTGGCGCCACCCCTGTGTCCCCCGTCTCGAG 180

Db 377 CCAGAGGAGCGCCCTAACGCGGCATCTGCTGGCGCCACCCCTGTGTCCCCCGTCTCGAG 436

Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGAGTACGCTCTGGACCTTCAGACATTTCTCCAG 240

Db 437 CCCTTTAATCTGACTTCGGAGCCCTCGAGTACGCTCTGGACCTTCAGACATTTCTCCAG 496

Qy 241 CAACACCGCGCGCTTC 258

Db 497 CAACACCGCGCGCTTC 514

RESULT

5

BE621346

LOCUS

DEFINITION BE621346 762 bp mRNA EST
 601493736F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3895736 5',
 mRNA sequence.

```

ACCESSION BE621346
VERSION BE621346.1 GI:9892286
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
JOURNAL 1 (bases 1 to 762)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM674 row: h column: 09
High quality sequence stop: 613.
Location/Qualifiers
1..762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3895736"
/clone_lib="NIH-MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site:1; Noli;
Site:2; Sali; Cloned unidirectionally. Primer: Oligo df.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 155 a 288 c 193 g 126 t
ORIGIN

Query Match 98.6%; Score 254.4; DB J10; Length 762;
Best Local Similarity 99.6%; Pred. No. 8.9e-55;
Matches 255; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCCGGCCCGCCAGTGCAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
DB 209 CTCGGCCCGCCAGTGCAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 268
QY 61 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCCTGCCT 120
DB 269 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCCTGCCT 328
QY 121 CCAGAGGACGCCCTTAACGGCCGATCCCTGGCGCCACCCCTGTCCTCCCGTCCGAG 180
DB 329 CCAGAGGACGCCCTTAACGGCCGATCCCTGGCGCCACCCCTGTCCTCCCGTCCGAG 388
QY 181 CCCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 240
DB 389 CCCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 448
QY 241 CAACACCCGGCGCT 256
DB 449 CAACACCCGGCGCTT 464

RESULT 6
LOCUS BE882130 916 bp mRNA EST 27-SEP-2000
DEFINITION 601505227F2 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906823 5',
mRNA sequence.
ACCESSION BE882130
VERSION BE882130
KEYWORDS EST.
SOURCE human.

ACCESSION BE273386
VERSION BE273386
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
JOURNAL 1 (bases 1 to 603)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM703 row: f column: 08
High quality sequence stop: 659.
Location/Qualifiers
1..916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3906823"
/clone_lib="NIH-MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site:1; Noli;
Site:2; Sali; Cloned unidirectionally. Primer: Oligo df.
Average insert size 2.1 kb.
BASE COUNT 187 a 327 c 268 g 134 t
ORIGIN

Query Match 96.7%; Score 249.4; DB J37; Length 916;
Best Local Similarity 99.6%; Pred. No. 1.7e-53;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCCGGCCCGCCAGTGCAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
DB 180 GTCCGGCCCGCCAGTGCAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 239
QY 61 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCCTGCCT 120
DB 240 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCCTGCCT 299
QY 121 CCAGAGGACGCCCTTAACGGCCGATCCCTGGCGCCACCCCTGTCCTCCCGTCCGAG 180
DB 300 CCAGAGGACGCCCTTAACGGCCGATCCCTGGCGCCACCCCTGTCCTCCCGTCCGAG 359
QY 181 CCCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 240
DB 360 CCCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 419
QY 241 CAACACCCGGCGCT 251
DB 420 CAACACCCGGCGCT 430

RESULT 7
LOCUS BE273386
DEFINITION 601143493F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:3507191 5',
mRNA sequence.
ACCESSION BE273386
VERSION BE273386
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
JOURNAL 1 (bases 1 to 603)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM703 row: f column: 08
High quality sequence stop: 659.
Location/Qualifiers
1..916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3906823"
/clone_lib="NIH-MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site:1; Noli;
Site:2; Sali; Cloned unidirectionally. Primer: Oligo df.
Average insert size 2.1 kb.
BASE COUNT 187 a 327 c 268 g 134 t
ORIGIN

Query Match 96.7%; Score 249.4; DB J37; Length 916;
Best Local Similarity 99.6%; Pred. No. 1.7e-53;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCCGGCCCGCCAGTGCAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
DB 180 GTCCGGCCCGCCAGTGCAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 239
QY 61 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCCTGCCT 120
DB 240 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCCTGCCT 299
QY 121 CCAGAGGACGCCCTTAACGGCCGATCCCTGGCGCCACCCCTGTCCTCCCGTCCGAG 180
DB 300 CCAGAGGACGCCCTTAACGGCCGATCCCTGGCGCCACCCCTGTCCTCCCGTCCGAG 359
QY 181 CCCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 240
DB 360 CCCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 419
QY 241 CAACACCCGGCGCT 251
DB 420 CAACACCCGGCGCT 430

RESULT 7
LOCUS BE273386
DEFINITION 601143493F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:3507191 5',
mRNA sequence.
ACCESSION BE273386
VERSION BE273386
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
JOURNAL 1 (bases 1 to 603)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM674 row: h column: 09
High quality sequence stop: 613.
Location/Qualifiers
1..762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3895736"
/clone_lib="NIH-MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site:1; Noli;
Site:2; Sali; Cloned unidirectionally. Primer: Oligo df.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 155 a 288 c 193 g 126 t
ORIGIN

Query Match 98.6%; Score 254.4; DB J10; Length 762;
Best Local Similarity 99.6%; Pred. No. 8.9e-55;
Matches 255; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCCGGCCCGCCAGTGCAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
DB 209 CTCGGCCCGCCAGTGCAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 268
QY 61 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCCTGCCT 120
DB 269 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCCTGCCT 328
QY 121 CCAGAGGACGCCCTTAACGGCCGATCCCTGGCGCCACCCCTGTCCTCCCGTCCGAG 180
DB 329 CCAGAGGACGCCCTTAACGGCCGATCCCTGGCGCCACCCCTGTCCTCCCGTCCGAG 388
QY 181 CCCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 240
DB 389 CCCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 448
QY 241 CAACACCCGGCGCT 256
DB 449 CAACACCCGGCGCTT 464

RESULT 6
LOCUS BE882130 916 bp mRNA EST 27-SEP-2000
DEFINITION 601505227F2 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906823 5',
mRNA sequence.
ACCESSION BE882130
VERSION BE882130
KEYWORDS EST.
SOURCE human.

```

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LHCMI86 row: j column: 24
High quality sequence stop: 579.
Location/Qualifiers

FEATURES

source
1..603
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3507191"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 112 a 241 c 154 g 96 t
ORIGIN

Query Match 95.1%; Score 245.4; DB 106; Length 603;
Best Local Similarity 99.2%; Pred. No. 1.7e-52;
Matches 257; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1 GTCCGGCGCCAGCTCCAGTCGAGAACCGAACCCACCCAAAGGCTTCTTCTCTGCTG 60
Db 177 GTCCGGCGCCAGCTCCAGTCGAGAACCGAACCCACCCAAAGGCTTCTTCTCTGCTG 236
Oy 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 120
Db 237 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 296
Oy 121 CCAGAGACGCCCTAAGCGGATCCTGTGCGCCACCCCTGTGTCC-CCGTCCTCGA 179
Db 297 CCAGAGACGCCCTAAGCGGATCCTGTGCGCCACCCCTGTGTCC-CCGTCCTCGA 356
Oy 180 CCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCCTCCA 239
Db 357 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCCTCCA 416
Oy 240 GCAACACCCGGCCGCTTC 258
Db 417 GCAACACCCGGCCGCTTC 435

RESULT 0
AA410666 463 bp mRNA EST 18-MAY-1997
LOCUS zt30g09.r1 Soares ovary tumor NBH07 Homo sapiens cDNA clone
DEFINITION IMAGE:723904 5' similar to SW:GL96_MOUSE P46694 IMMEDIATE EARLY
PROTEIN GLY96. ; mRNA sequence.
AA410666
AA410666.1 GI:2069789
EST.
Homo sapiens
ORGANISM

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, H., Schellenberg, K., Steptoe, M., Tan, F., Theising, R., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL. contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 Et from Amersham
High quality sequence stop: 369.
Location/Qualifiers

FEATURES

source
1..463
/organism="Homo sapiens"
/db_xref="GDB:5935321"
/db_xref="taxon:9606"
/clone="IMAGE:723904"
/clone_lib="Soares ovary tumor NBH07"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT713D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTCGAGCGCGGCTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT713 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 71 a 196 c 114 g 82 t
ORIGIN

Query Match 94.8%; Score 244.6; DB 7; Length 463;
Best Local Similarity 98.4%; Pred. No. 2.6e-52;
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GTCCGGCGCCAGCTCCAGTCGAGAACCGAACCCACCCAAAGGCTTCTTCTCTGCTG 60
Db 213 GTCCGGCGCCAGCTCCAGTCGAGAACCGAACCCACCCAAAGGCTTCTTCTCTGCTG 272
Oy 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 120
Db 273 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 332
Oy 121 CCAGAGACGCCCTAAGCGGATCCTGTGCGCCACCCCTGTGTCC-CCGTCCTCGA 180
Db 333 CCAGAGACGCCCTAAGCGGATCCTGTGCGCCACCCCTGTGTCC-CCGTCCTCGA 392
Oy 181 CCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCCTCCA 240
Db 393 CCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCCTCCA 452
Oy 241 CAACACCCGGC 251
Db 453 CAACACCCGGC 463

RESULT 9
AW960290 585 bp mRNA EST 01-JUN-2000
LOCUS EST372361 MAGE resequences, MAGF Homo sapiens cDNA, mRNA sequence.
DEFINITION AW960290
ACCESSION AW960290
VERSION AW960290.1 GI:8149974
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE: Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 585)
 AUTHORS: Hegde, P., Ol, R., Aherthy, K., Dharap, S., Gaspard, R., Gay, C., Holt
 I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J., and
 Quackenbush, J.

TITLE: Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray

JOURNAL: Unpublished (2000)
 COMMENT: Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 144

Seq primer: Reverse*

Location/Qualifiers

1..585
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGP"
 /note="Vector: pBluescriptSKm"

BASE COUNT 134 a 190 c 148 g 113 t

ORIGIN

Query Match 93.8%; Score 242; DB 97; Length 585;

Best Local Similarity 96.1%; Pred. No. 1.2e-51;

Matches 248; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GTCCGGCCCGCCAGCTGCCAGTCAGGAGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60

Db 19 GCGCGGCCAGCTGCCAGTCAGGAGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 78

QY 61 CTCACCATGCTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCGGCGCCCTGCT 120

Db 79 CTCACCATGCTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCGGCGCCCTGCT 138

QY 121 CCAGAGGAGCGCCCTAACCGCGCATCCCTGCGGCCACCCCTGTGTCGCCCGTCTCTCAG 180

Db 139 CCAGAGGAGCGCCCTAACCGCGCATCCTGAGGCCACCCCTGTGTCGCCCGTCTCTCAG 198

QY 181 CCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTCGACCTCAGACACTTCTCTCCAG 240

Db 199 CCTTTAATCTGACTTCGAGCCCTCGGACTACGCTCTCGACCTCAGACACTTCTCTCCAG 258

QY 241 CAACACCGCGCGCCTTC 258

Db 259 CAACACCGCGCGCCTTC 276

RESULT 10

BE563592

LOCUS

DEFINITION BE563592 810 bp mRNA EST 15-AUG-2000

601334706P1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3688825 5',

mRNA sequence.

ACCESSION BE563592

VERSION BE563592.1 GI:9807312

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 810)

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LCM382 row: k column: 02

High quality sequence stop: 692.

FEATURES

source

1..810

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:3688825"

/tissue_type="adenoecarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;

Site: 2: EcoRI; cDNA made by oligo-dT priming

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAC(G). Library constructed

by Ling Hong in the laboratory of Gerald M. Rubin

(University of California, Berkeley) using ZAP-cDNA

synthesis kit (Stratagene) and Superscript II RT (Life

Technologies)."

BASE COUNT 176 a 284 c 205 g 145 t

ORIGIN

Query Match 91.8%; Score 236.8; DB 109; Length 810;

Best Local Similarity 98.8%; Pred. No. 2.6e-50;

Matches 249; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GTCCGGCCCGCCAGCTGCCAGTCGAGGAGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60

Db 207 GTCCGGCCCGCCAGCTGCCAGTCGAGGAGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 266

QY 61 CTCACCATGCTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCGGCGCCCTGCT 120

Db 267 CTCACCATGCTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCGGCGCCCTGCT 326

QY 121 CCAGAGGAGCGCCCTAACCGCGCATCCCTGCGGCCACCCCTGTGTCGCCCGTCTCTCAG 179

Db 327 CCAGAGGAGCGCCCTAACCGCGCATCCCTGCGGCCACCCCTGTGTCGCCCGTCTCTCAG 386

QY 180 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCA 239

Db 387 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCA 446

QY 240 GCAACACCGGC 251

Db 447 GCAACACCGGC 458

RESULT 11

W52268

LOCUS

DEFINITION W52268 447 bp mRNA EST 11-OCT-1996

zc46a10.r1 Soares_senescent_fibroblasts_NBHSF Homo sapiens cDNA

clone IMAGE:325338 5' similar to SW:GL96_MOUSE P46694 IMMEDIATE

EARLY PROTEIN GLY96. [1]; mRNA sequence.

ACCESSION W52268

VERSION W52268.1 GI:1349380

KEYWORDS EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 447)

Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Paterson, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston

R., Williamson, A., Wohlmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewartson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1695 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 392.

Location/Qualifiers

FEATURES

source

1..504

/organism="Homo sapiens"

/db_xref="GDB:4623266"

/db_xref="taxon:9606"

/clone="IMAGE:591880"

/clone_lib="Stratagene pancreas (#937208)"

/lab_host="SOLR cells (kanamycin resistant)"

/notes="Organ: pancreas; Vector: pBluescript SK-; Site:1:

ECORI; Site:2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Pancreatic adenocarcinoma cell line. Average

insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor

sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 80 a 202 c 121 g 92 t 9 others

ORIGIN

Query Match

Best local Similarity 76.7%; Score 197.8; DB 2; Length 504;

Matches 233; Conservative 0; Mismatches 19; Indels 5; Gaps 2;

```

Qy 1  GTCCGGCGGCGGAGTCCGAGTCCGAGAACCC-AGCCAAAGGCTTCTCTTTCTGCT 59
    |||||||  |  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 225 GTCCGGCGGCGGAGTCCGAGTCCGAGAACCC-AGCCAAAGGCTTCTCTTTCTGCT 284
    |||||||  |  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Qy 60  GCTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGCGCCCTGCC 119
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 285 GCTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGCGCCCTGCC 344
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Qy 120 TCCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGGTCCTCGA 179
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 345 TCCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGGTCCTCGA 404
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Qy 180 GCCCTTTAATCTGACTTCGGAGCCCTCGG----ACTACGCTCTGGACCTCAGCACTTCC 235
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 405 GCNCTTTAATCTGACTTCGGAGCCCTTGGAACTANGCTTCTTGGAACTTAAGCACTTCC 464
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Qy 236 TCCAGCAACACCCGGCC 252
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 465 TCCAGCAACACCCGGCC 481
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

Search completed: January 30, 2001, 18:47:28

Job time: 28459 sec

by RNA polymerase II, the RNA polymerase which transcribes messenger RNA. RNA polymerase II transcription proceeds *in vitro* upon addition of several nuclear fractions designated TP1A, B, D, E, H, I and J to RNA polymerase II holoenzyme. Fraction TF1H has been shown to contain a TBP and other TAFs. Purification of h1ID and separation of its components reveals 7 proteins ranging in size from 30-250 kD. Serum raised against the TF1D fraction allowed cloning of the cDNAs from lambda-gt10 expression libraries.

SQ Sequence 1120 BP; 307 A; 291 C; 312 G; 210 T; 0 other;

Query Match 13.0%; Score 33.6; DB 17; Length 1120;
Best Local Similarity 51.0%; Pred. No. 2.2;
Matches 104; Conservative 0; Mismatches 99; Indels 1;

[illegible]

RESULT 15

T79601/c
ID T79601 standard; cDNA: 1120 BP.

AC T79601:

XX	08-OCT-1997	(first entry)
DT		

DE TATA-binding protein associated factor, dTAFII40 cDNA.

TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP;
 KW initiation; ds.

OS *Drosophila* sp.

AA	FH	Key	Location/Qualifiers
FT	.	CDS	80..916
FT			/*tag= a

XX
PN
US5637686-A.

10-JUN-1997.

28-JAN-1993; 93US-0013412.

28-JAN-1994; 94US-0188582.

PR 28-JAN-1993; 93US-0013412;
PR 30-JUN-1993; 93US-0087119;

PK 09-MAY-1990; 9605-0646/13:
XX

PA (REGC) UNIV CALIFORNIA:
XX

PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;

PI Tjian R, Wang E, Weinzierl ROJ;

DR WPI: 1997-319113/29.

DR P-PSDB; W25025.

PT Nucleic acids encoding human TATA-binding protein associated factor

(TAF) peptide(s) - for production of recombinant peptide(s), used for modulating transcription of TAFs

Example 1; Column 61-64; 86pp; English.

T79601 encodes Drosophila TATA-binding protein associated factor (TAF) polypeptide, dTAFII40 (mol. weight 40kD). TAF peptides derived from dTAFII30 alpha, dTAFII30 beta, dTAFII40, dTAFII60, dTAFII80, dTAFII10, dTAFII50, and dTAFII250, their human equivalents and nucleic acids encoding them, are used to modulate transcription, including transcription initiation. TAFs are nuclear proteins involved in RNA polymerase I, II and III transcription. The peptides act by binding to a different TAF, an activator, or TBP (TATA-binding protein) or competitively inhibiting association of a TAF domain with another compound, typically a protein like TBP or another TAF, an activator, or DNA.

Sequence 1120 BP; 307 A; 291 C; 312 G; 210 T; 0 other;

Query Match	13.0%	Score 33.6;	DB 18;	Length 1120;
Best Local Similarity	51.0%;	Pred. No. 2.2;		
Matches 104;	Conservative	0;	Mismatches 99;	Indels 1;
Gaps 1;				

[illegible]

Search completed: January 30, 2001, 21:29:32
Job time: 19258 sec

CC Surface molecule 4F2. Antisense DNA or RNA, antibodies, antibody
 CC fragments, and peptide and non-peptide antagonists to LAT1 are useful
 CC as anticancer agents. The present sequence encodes human LAT1, which is
 CC specifically claimed in the present invention.
 XX
 SQ Sequence 4539 BP; 765 A; 1450 C; 1297 G; 998 T; 29 other;

Query Match 13.1%; Score 33.8; DB 21; Length 4539;
 Best Local Similarity 53.0%; Pred. No. 2.6;
 Matches 62; Conservative 2; Mismatches 53; Indels 0; Gaps 0;
 QY 8 GCCAGCTGCCAGTCGAGGACGACGACCCAGCCAAAGGCTTCCTTCCTGCTCAACCA 67
 DB 556 gccgggtgccggaggagcagcagcagctgtggcctggytntggtgtgtgtcagcg 615
 QY 68 TCCCTCTCCACACCTCAGCTCAAGACGCTCTCCCGCCCTCCCTCCACG 124
 DB 616 ccylgaactgclacagcgtgaaggcngcnacncggatccaggatgcltgcgcgcg 672

RESULT 13
 Q70728/C
 ID Q70728 standard; cDNA; 1120 BP.

XX Q70728;

XX 23-MAR-1995 (first entry)

XX TATA-binding protein-associated factor dTAFII40 cDNA.

DE TATA-binding protein associated factor; dTAFII40; ss; screening;
 KW diagnostic; therapeutic; gene transcription regulation.

XX Drosophila.

XX Key Location/Qualifiers
 FH 80..913
 FT CDS /*tag= a

XX W09417087-A.

XX 04-AUG-1994.

XX 28-JAN-1994; 94WO-US01114.

XX 28-JAN-1993; 93US-0013412.

XX 30-JUN-1993; 93US-0087119.

XX (REGC) UNIV CALIFORNIA.

XX Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;

XX WPI: 1994-264019/32.

XX P-PSDB; R56490.

XX TATA-binding protein associated protein factors - and
 PT corresponding nucleotide sequence and deriv. antibodies, useful
 PT in screening, diagnostics and therapeutics

XX Disclosure; Page 89-92; 180pp; English.

XX The TATA-binding protein associated factor dTAFII40 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.

XX Sequence 1120 BP; 307 A; 292 C; 311 G; 210 T; 0 other;

Query Match 13.0%; Score 33.6; DB 15; Length 1120;
 Best Local Similarity 51.0%; Pred. No. 2.2;
 Matches 104; Conservative 0; Mismatches 99; Indels 1; Gaps 1;
 QY 34 CCAGCCAAAGGCTTCTCTTCTGCTCACCATCGTCTTCTGCCAGATCTGATGGCT 93
 DB 866 CCACTCCAAGCCAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 808
 QY 94 GAAGAGGTGTCCGGCGCCCTGCTCCAGAGGAGCCCTTAACCCGCATCCCTGGCG 153
 DB 807 GAACCGGATGCCCGCAGGATGCTGCCACGCGCGCGCATCGTGTCCACCTCCATCTTG 748
 QY 154 CCCACCCCTGTGTCCCGCTCTCGATCTCTTTAATCTGACTTGGAGCCCTGGACTAC 213
 DB 747 AGATCTCCCGCGCGCGCTGCTCTCGCTCTTAACCTCTCTGACCCAGCCGCCGCCAGAT 688
 QY 214 GCTCTGGACCTCAGCACCTTTCCTC 237
 DB 687 CGCGCGGAGCTGCCACCGTTTTC 664

RESULT 14
 T42214/C
 ID T42214 standard; cDNA; 1120 BP.

XX T42214;

XX 27-JAN-1997 (first entry)

XX Drosophila TATA-binding protein associated factor dTAFII60 gene.

DE Drosophila; TATA-binding protein; TBP associated factor; TFIIID;
 KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
 KW holoenzyme; lambda-gt11; expression library; ds.

XX Drosophila melanogaster.

XX Key Location/Qualifiers
 FH 80..916
 FT CDS /*tag= a

XX /product= Drosophila TAFII60

XX US5534410-A.

XX 09-JUL-1996.

XX 28-JAN-1993; 93US-0013412.

XX 28-JAN-1994; 94US-0188582.

XX 28-JAN-1993; 93US-0013412.

XX 30-JUN-1993; 93US-0087119.

XX (REGC) UNIV CALIFORNIA.

XX Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;

XX WPI: 1996-333245/33.

XX P-PSDB; T42214.

XX Screen for cpds. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.

XX Examples; Column 61-64; 86pp; English.

XX This is the nucleotide sequence encoding the Drosophila TATA-binding
 CC protein (TBP) associated factor (TAF) designated TAFII60. The protein
 CC is a component of the TFIIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The encoded protein
 CC has an estimated mol. wt. of 60 kD by SDS-PAGE.

XX The invention relates to purified proteins involved in transcription

	Query Match	14.08;	Score 36.2;	DB 21;	Length 1000;
	Best Local Similarity	44.72;	Pred. No. 0.42;		
	Matches 68;	Conservative	0;	Mismatches	84;
				Indels	0;
				Gaps	0;
QY	106	CCGCGCCCTGCTCCAGAGAGCGCCCTAACGCCGCATCCCTGGCGCCACCCCTGTG	165		
UB	803	CTTCCGCCCCCCCCNCGGNGNCCGNTCCCGCCNCTCCCGGNNCCGCGCCNCCN	744		

V69706 standard; DNA; 170 BP.
V69706:
08-FEB-1999 (first entry)
Nucleotide sequence of intron 3 of iENTP gene.
Equilibrative nucleoside transport protein; iENTP; NBMPK; transport;
nitrobenzylmercaptopyrine riboside; antiviral; antitumour; screening;
inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;
adenosine deaminase; factor VIII; intron; ds.
Homo sapiens.
WO9846749-A1.
22-OCT-1998.
10-APR-1998; 98WO-US07283.
09-APR-1998; 98US-0058389.
11-APR-1997; 97US-0838645.
(SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
Belt JA, Crawford CR, Patel DH;
WPI; 1998-594576/50.
New equilibrative nucleoside transport protein insensitive to
nitrobenzylthio-purine riboside - useful for, e.g. identifying
specific therapeutic nucleoside analogues and in gene therapy to
protect transduced cells against ablative chemotherapy
Disclosure; Page 88; 114pp; English.
Sequences V69704 to V69715 represent introns of the gene encoding an
equilibrative nucleoside transport protein (iENTP), which is insensitive
to nitrobenzylmercaptopyrine riboside (NBMPR). Cells transformed with a
construct containing the iENTP nucleic acid can be used to produce the
protein recombinantly. iENTP is used to identify specific ligands
(particularly antiviral and antitumour nucleoside analogues that are
preferentially transported into cells) and to raise antibodies. Cells in
which iENTP provides all available transport activity are used: (a) to
identify permeants of iENTP and (b) to screen specific inhibitors of
iENTP (potential drugs). Fragments of the iENTP nucleic acid are used,
as probes, primers, antisense molecules, and ribozymes for therapy or
diagnosis, and knockout mice in which both alleles encoding iENTP
contain an inactivating defect are also useful for drug screening. Cells
that have been transduced with iENTP nucleic acid ex vivo are used
particularly, for cancer chemotherapy. Vectors in which the iENTP gene
is linked to a heterologous gene (e.g. encoding adenosine deaminase and
factor VIII) can be used for gene therapy.
Sequence 170 BP; 18 A; 77 C; 42 G; 33 T; 0 other;

242235 standard; cDNA; 1526 BP.
242235:
31-JAN-2000 (first entry)
Human normal bladder tissue cDNA derived EST 114.
Human; bladder; treatment; EST; expressed sequence tag; cytosolic;
cancer; gene therapy; ss.
Homo sapiens.
DE19818620-A1.
28-OCT-1999.
21-APR-1998; 98DE-1018620.
21-APR-1998; 98DE-1018620.
(META-) METAGEN GES GENOMFORSCHUNG MBH.
Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
WPI; 1999-602416/52.
P-PSDB; Y60556, Y60557, Y60558.
New polypeptides and their nucleic acids, useful for treatment of
bladder tumour and identification of therapeutic agents
Claim 3; Page 234; 366pp; German.
This invention describes novel polypeptide fragment sequences (I) and
their encoding nucleic acids (II) which are highly expressed in normal
bladder tissue and have cytostatic activity. (II) are used for
recombinant expression of (I) and to isolate complete genes. (I) are
used to identify agents suitable for the treatment of bladder tumours, to
directly treat this form of cancer (including expression from gene
therapy vectors), or are used in a preparation for cancer treatment. (I)
is also used for the generation of specific antibodies. (II) are
identified by assembling ESTs (expressed sequence tags) from a particular
tissue type before comparison of expression patterns. This allows a
significantly longer fragment of the gene to be revealed, and therefore
reduces the number of failures because of ESTs from different libraries
representing different parts of the same unknown gene distorting the
estimated frequency of occurrence in a particular tissue 242122-242248
represent EST fragments derived from a human normal bladder tissue cDNA
library which encode the protein fragments represented in Y60329-Y60591.
Sequence 1526 BP; 365 A; 523 C; 437 G; 201 T; 0 other;

Query Match 14.1%; Score 36.4; DB 20; Length 1526;
Best Local Similarity 50.4%; Pred. No. 0.41;
Matches 114; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
Qy 2 TCGCGGCCAGCTGCCAGTCCGAGGACCGAACCCAGCCAAAGGCTTCTTTCTGCTGTC 61
Db 240 TGTCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
Qy 62 TCACCATCGTCTTCTGCGCAGATCCGTGATGCTGAAGAGGTGTCGCGGCCCTCTGCTC 121
Db 180 TGTGCTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 121
Qy 122 CAGAGGACGCCCTTACCGCGCATCCCTGGCGGCCACCCCTGTGTCCTCCCTCTCTGAGC 181
Db 121 CCCTACACGCTCAAGTCCGCCCCACGGATGGCGGCCGCCGCCGCCGCCGCCGCCGCC 62
Qy 182 CTTTATTCGATTCGGAGCCCTTCGAGCTACGCTCTGGACCTCAG 227
Db 61 GCGCGGCTCGCGGCCCGGGGCTCCCGGCTCCCGGCTCCCGGCTGCTACTCCG 16

Query Match 14.6%; Score 37.6; DB 19; Length 170;
Best Local Similarity 65.5%; Pred. No. 0.12; Indels 29; Gaps 0;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
Qy 101 GTGTGCGCGGCCCTGCTCCAGAGGACGCCCTAACCGCGCATCCCTGCGGCCACCC 160
Db 1 gtgagagccctgcccctgctcgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 60
Qy 161 CTGTGTCCTCCCTCTCGAGCCCT 184
Db 61 ctgagcccccctgcccctccagccct 84

PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Belt JA, Crawford CR, Patel DH;

WPI; 1998-594576/50.

PT New equilibrative nucleoside transport protein insensitive to
PT nitrobenzylthio-purine riboside - useful for, e.g. identifying
PT specific therapeutic nucleoside analogues and in gene therapy to
PT protect transduced cells against ablative chemotherapy

PS Example 3; Pages 76-79; 114pp; English.

This represents the 5' UTR (untranslated region) of the DNA that encodes an equilibrative nucleoside transport protein (iENTP), which is insensitive to nitrobenzylmercaptapurine riboside (NBMPR). Cells transformed with a construct containing the iENTP nucleic acid can be used to produce the protein recombinantly. iENTP is used to identify specific ligands (particularly antiviral and antitumour nucleoside analogues that are preferentially transported into cells) and to raise antibodies. Cells in which iENTP provides all available transport activity are used: (a) to identify permeants of iENTP and (b) to screen specific inhibitors of iENTP (potential drugs). Fragments of the iENTP nucleic acid are used, as probes, primers, antisense molecules, and ribozymes for therapy or diagnosis, and knockout mice in which both alleles encoding iENTP contain an inactivating defect are also useful for drug screening. Cells that have been transduced with iENTP nucleic acid *ex vivo* are used particularly, for cancer chemotherapy. Vectors in which the iENTP gene is linked to a heterologous gene (e.g. encoding adenosine deaminase and factorVIII) can be used for gene therapy.

Sequence 6354 BP; 1322 A; 1797 C; 1873 G; 1355 T; 7 other;

Query Match 15.0%; Score 38.6; DB 19; Length 6354;
Best Local Similarity 65.9%; Pred. No. 0.14;
Matches 56; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 100 GGTGTGCGCGCGCCCTGCCCTCCAGAGGACGCCCTAAGCGGCATCCCTGGCGCCACC 159

Db 2461 ggtagagagcctgccctggctgacctgcgcctctgcgagggcagctctcattgagggcctcc 2520

QY 160 CCTGTGTCCTCGAGCCCCT 184

Db 2521 cctgcgccccctgcctccagccct 2545

RESULT 6

251683

IV 251683 standard; cDNA; 4228 bp.

AC 251683:

04-JUL-2000 (first entry)

Human cyclic nucleotide-associated protein-2 (CNAP-2) cDNA.

Cyclic nucleotide-associated protein-2; CNRP-2; human; cytosolic; anti-arteriosclerotic; hepatotropic; anti-leukemic; anti-inflammatory; immunomodulatory; anti-asthmatic; anti-anemic; anti-diabetic; diagnosis; anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer; anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological; anti-infertility; anti-allergic; vasotropic; immunosuppressive; hypertensive; gene therapy; prevention; treatment; arteriosclerosis; cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus; neurological; vision; reproductive; smooth muscle; ss.

OS Homo sapiens.

Key	Location/Qualifiers
-----	---------------------

FT CDS 31..4014

```

F11
F11
/*tag= a
/product= "Human CNAP-2 protein"

```

FT	/note= "Shares 24% identity to Aquifex pyrophilus"
FT	esterase 28LC"
FT	31..132
FT	/*tag= b
FT	133..4011
FT	/*tag= c
FT	/product= "Mature CNAP-2 protein"
FT	136..165
FT	/*tag= d
FT	/bound_motley= "Primer or probe"
FT	/note= "Useful for amplification or hybridisation techniques"
FT	

PN WO200014248-A1.

16-MAR-2000.

AA
PF 03-SEP-1999; 99WO-US20287.

PR 04-SEP-1998; 98US-0148904.

PA (INCY-) INCYTE PHARM INC.

PI	Hillman JL,	Yue H,	Guegler KJ,	Corley NC,	Patterson C,	"Tang Y":
----	-------------	--------	-------------	------------	--------------	-----------

DR WPI; 2000-256994/22.

DR P-PSDB; Y70474.

Isolated cyclic nucleotide associated proteins useful for preventing, diagnosing and treating cell proliferative, autoimmune/inflammatory, neurological, vision, reproductive and smooth muscle disorders -

PS Claim 9; Page 70-71; 78pp; English.

The present sequence is the cDNA encoding human cyclic nucleotide associated protein-2 (CNAP-2), identified in Incyte clone 3149674, that is isolated from ADRENOM4 cDNA library. It is expressed in nervous, reproductive, cardiovascular and haematopoietic/immune tissues. CNAP sequences may be used for prevention, treatment and diagnosis of diseases associated with altered CNAP expression such as, cell proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia, lymphoma and cancer of the breast, prostate, lung and brain), inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple sclerosis and psoriasis), neurological disorders (e.g. epilepsy, Alzheimer's/parkinson's disease and strokes), vision disorders (e.g. conjunctivitis, glaucoma, cataracts and retinitis pigmentosa), reproductive disorders (e.g. infertility, uterine fibroids, ectopic pregnancies and impotence) and smooth muscle disorders (e.g. angina, anaphylactic shock, Kearns-Sayre syndrome and hypertension). It can also be used for gene therapy.

Sequence 4228 BP; 793 A; 1365 C; 1324 G; 746 T; 0 other;

Query Match	14.8%	Score 38.2;	DB 21;	Length 4228;
Best Local Similarity	54.7%	Pred. No. 0.17;		
Matches 76;	Conservative	0;	Mismatches 63;	Indels 0;
Matches				Gaps 0;

QY 75 CTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCCTCCAGAGGACGCCCC 134
||| || | |||| | |||| | | | | | | | |
Db 1185 ctccgaactcgacatggcctatgagcgtgcgcgatctccctgtccctgcagaagaagtc 1244

QY
135 TAACGGCGCATCCTTGGGCCCCACCCCTGTGTCCCGCTCTCGAGTCCTTTAATCTCAC 194

Db
1245 ctccgggggtctcttgagccccgcctgagaccccactcaggagacctctgaagcagacc 1304

QV 195 TTCGGAGCCCTCGGACTAC 213

Db 1305 qccagccctatgaatac 1323

RESULT
V69706
7

```
DT 15-FEB-1999 (first entry)
XX EST clone BY66.
XX
XX Human: secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
XX WO9845436-A2.
XX
XX 15-OCT-1998.
XX
XX 10-APR-1998: 98WO-US06955.
XX
XX 10-APR-1997: 97US-0838821.
XX
XX (CEMY ) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX Racie LA, Spaulding V, Treacy M;
XX
XX WPI: 1999-070077/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pituitary, retina and colon cDNA libraries.
XX
XX Claim 1: Page 164; 618pp; English.
XX
XX The present sequence represents a human expressed sequence tag (EST).
XX The polynucleotide, which is a secreted EST, and the encoded protein
XX are predicted to have useful biological activities which would make
XX them suitable for treating, preventing or ameliorating medical
XX conditions in humans and animals, although no supporting data is
XX given. Suggested activities include nutritional activity, immune
XX stimulating or suppressing activity, haematopoiesis regulating
XX activity, tissue growth activity, activin/inhibin activity,
XX chemotactic/chemokinetic activity, haemostatic and thrombolytic
XX activity, receptor/ligand activity, anti-inflammatory activity,
XX cadherin/tumour invasion suppressor activity, tumour inhibition
XX activity. The polynucleotide may also be useful for gene therapy.
XX
XX Sequence 193 BP; 35 A; 39 C; 58 G; 51 T; 0 other;
XX
XX
XX Query Match 32.6%; Score 84; DB 20; Length 193;
XX Best Local Similarity 100.0%; Pred. No. 4.2e-14;
XX Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 175 CTCGAGCCCTTTAACTGACTTCGGAGCCCTCGGACTACGCTCTCGACCTCAGCATTTC 234
XX |||||||
XX DB 193 CTCGAGCCCTTTAACTGACTTCGGAGCCCTCGGACTACGCTCTCGACCTCAGCATTTC 134
XX
XX QY 235 CTCGAGCAACACCCGCGCCCTTC 258
XX |||||||
XX DB 133 CTCGAGCAACACCCGCGCCCTTC 110
XX
XX RESULT 4
XX Q77534
XX ID Q77534 standard; DNA: 297 BP.
XX
XX AC Q77534;
XX
XX DT 23-SEP-1994 (first entry)
XX
XX DE Human genome fragment.
XX
XX KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
XX detection; homology; human; adrenal tissue; ds.
```

```
XX Homo sapiens.
OS
XX WO9401548-A.
PN
XX 20-JAN-1994.
PD
XX 13-JUL-1993: 93WO-GB01467.
PF
XX 13-JUL-1992: 92GB-0014857.
PR
XX (MEDI-) MEDICAL RES COUNCIL.
PA
XX Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
XX Sibson DR, Starkey M;
PI
XX WPI: 1994-035056/04.
DR
XX New nucleic acid fragment encoding gene products - can be used
XX for genetic analysis and mapping
PT
XX Claim 1: Page 575-576; 616pp; English.
PS
XX Human nucleic acid fragments, isolated from brain, adrenal tissue,
XX the placenta or bone marrow comprise any of: (A) a sequence
XX selected from (Q76401-Q77613), (B) an allelic variation of a
XX sequence as described in (A), or (C) a sequence complementary
XX to (A) or (B).
CC
XX Preferred sequences exhibit no more than 90% homology to a human
XX sequence known per se.
CC
XX Sequence 297 BP; 66 A; 117 C; 61 G; 52 T; 1 other;
XX
XX
XX Query Match 17.1%; Score 44; DB 15; Length 297;
XX Best Local Similarity 77.9%; Pred. No. 0.0026;
XX Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
XX
XX QY 2 TCCGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTGC 61
XX |||||||
XX DB 225 tccgcgccagctgccaaggaaggaacccagcagagctcttcttcggttga 284
XX |||||||
XX QY 62 TCACCATC 69
XX |||||||
XX DB 285 tcgccatc 292
XX
XX RESULT 5
XX V69699
XX ID V69699 standard; DNA: 6354 BP.
XX
XX AC V69699;
XX
XX DT 08-FEB-1999 (first entry)
XX
XX DE 5' UTR sequence of iENTP DNA.
XX
XX KW Equilibrative nucleoside transport protein; iENTP; NBMPR; transport;
XX nitrobenzylmercaptopyrine riboside; antiviral; antitumour; screening;
XX inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;
XX adenosine deaminase; factor VIII; UTR; untranslated region; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO9846749-A1.
XX
XX PD 22-OCT-1998.
XX
XX PF 10-APR-1998: 98WO-US07283.
XX
XX PR 09-APR-1998: 98US-0058389.
XX
XX PR 11-APR-1997: 97US-0838845.
XX
```

```

DR P-PSDB: W36006.
XX
XX New genes differentially expressed in cardiovascular disease - used
PT for diagnosis, drug screening and treatment of cardiovascular
PT disease, e.g. atherosclerosis, restenosis, hypertension, etc
XX
XX
XX Claim 1: Fig 5: 163pp: English.
XX
XX Fchd605 is a novel human gene that is up-regulated in monocytes
XX treated with oxidised low density lipoproteins (LDL). Differential
XX display was used to detect genes that are differentially expressed
XX in monocytes treated so as to simulate the conditions under which
XX foam cells develop during atherogenesis. Both fchd605 and fchd602
XX (see T94470) are up-regulated under the disease condition of
XX treatment with oxidised LDL. The fchd605 gene product (see W36006)
XX has sequence similarity to mouse gly96. The discovery of the
XX up-regulation of these 2 genes provides a fingerprint profile,
XX e.g. markers for the study of cardiovascular diseases, including
XX atherosclerosis, ischaemia/reperfusion, hypertension, restenosis
XX and arterial inflammation. Methods are provided for the diagnosis,
XX monitoring in clinical trials, screening for therapeutically
XX effective compounds, and treatment of cardiovascular diseases based
XX on discoveries regarding the expression patterns of novel genes
XX fchd531 (see T94467), fchd540 (see T94468), fchd545 (see T94469),
XX fchd602 and fchd605.
XX
XX Sequence 1228 BP; 265 A; 362 C; 340 G; 261 T; 0 other;

Query Match 100.0%; Score 258; DB 18; Length 1228;
Best Local Similarity 100.0%; Pred. No. 1e-60;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGCGCCAGCTCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCCTTCCTCG 60
Db 211 gtccgcgccagctgcagtcgaggaacccagcaacccagcaaaaggctctcttctg 270
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGCGCCCTGCCT 120
Db 271 ctaccatctcttctgccagatcctgattggtgaagaggtgtgccgccccctgct 330
Qy 121 CCAGAGAGCGCCCTAAGCGCGCATCCCTGGCGCCACCCCTGTGTGCCGCTCTCGAG 180
Db 331 ccagaggagcgccctaaacgcgcgcatccctggcgccaccctgtgtccccctcgag 390
Qy 241 CAACACCGCGCGCCTTC 258
Db 451 caacacccggcgcccttc 468

RESULT 2
ID 250711 standard; DNA; 1228 BP.
XX
XX AC 250711;
XX
XX 31-MAY-2000 (first entry)
XX
XX Nucleotide sequence of human fchd605 gene.
XX
XX fchd605 gene; human; cardiovascular disease; oncogenic disorder;
XX diabetic retinopathy; fibroproliferative disorder; atherosclerosis;
XX TGF-beta signalling pathway; TGF-beta transforming growth factor;
XX pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;
XX vascularisation; cytostatic; antidiabetic; ophthalmological; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers

```

```

CDS 1..471
FT /*tag= a
FT /product= "fchd605 protein"
XX
XX WO200006206-A1.
XX
XX 10-FEB-2000.
XX
XX 30-JUL-1999; 99WO-US17394.
XX
XX 30-JUL-1998; 98US-0126640.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Falb DA;
XX
XX WPI; 2000-205414/18.
XX
XX P-PSDB; Y45017.
XX
XX Identifying substances for ameliorating symptoms of fibroproliferative
XX diseases or oncogenic related disorders -
XX
XX Examples: Fig 5: 214pp; English.
XX
XX The patent discloses methods for the treatment and diagnosis of
XX cardiovascular diseases by novel human genes which are differentially
XX expressed in different cardiovascular disease states. Compositions which
XX can modify TGF-beta signalling pathway are identified by screening.
XX These are used therapeutically to treat fibroproliferative and oncogenic
XX disorders, especially TGF (Transforming growth factor)-beta related
XX disorders, including diabetic retinopathy, atherosclerosis, pancreatic
XX cancer, angiogenesis, inflammation, fibrosis, tumour growth and
XX vascularisation. The present sequence is fchd605 gene which is up
XX -regulated in monocytes treated with oxidised LDL (low density
XX lipoprotein) can be used to design cardiovascular disease treatment
XX strategies. Depending on whether the up-regulation has a pathogenic or
XX protective effect treatment methods can be designed to increase or
XX decrease the activity of the protein product of the gene.
XX
XX Sequence 1228 BP; 265 A; 361 C; 341 G; 261 T; 0 other;

Query Match 100.0%; Score 258; DB 21; Length 1228;
Best Local Similarity 100.0%; Pred. No. 1e-60;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGCGCCAGCTCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCCTTCCTCG 60
Db 211 gtccgcgccagctgcagtcgaggaacccagcaacccagcaaaaggctctcttctg 270
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGCGCCCTGCCT 120
Db 271 ctaccatctcttctgccagatcctgattggtgaagaggtgtgccgccccctgct 330
Qy 121 CCAGAGAGCGCCCTAAGCGCGCATCCCTGGCGCCACCCCTGTGTGCCGCTCTCGAG 180
Db 331 ccagaggagcgccctaaacgcgcgcatccctggcgccaccctgtgtccccctcgag 390
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCACACCTTCCTCCAG 240
Db 391 ccctttaatctgacttcggagccctcgactaogctctgagacctcagcactttctccag 450
Qy 241 CAACACCGCGCGCCTTC 258
Db 451 caacacccggcgcccttc 468

RESULT 3
ID V89271/c
XX V89271 standard; cDNA; 193 BP.
XX
XX AC V89271;
XX

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 21:29:29 ; Search time 77.57 seconds

(without alignments)

1249.465 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468

Perfect score: 258

Sequence: 1 GTCCGGCGCCAGTCCAGT.....AGCAACACCGCGCCGCTTC 258

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: -480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

```

1: /cgn2_2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /cgn2_2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /cgn2_2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /cgn2_2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /cgn2_2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /cgn2_2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /cgn2_2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /cgn2_2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /cgn2_2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /cgn2_2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /cgn2_2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /cgn2_2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /cgn2_2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /cgn2_2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /cgn2_2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /cgn2_2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /cgn2_2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /cgn2_2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	258	100.0	1228	18 T94471	Human Fchd605 gene
2	258	100.0	1228	21 T50711	Nucleotide sequenc
3	84	32.6	193	20 V89271	EST clone Bf66. H
4	44	17.1	297	15 Q77534	Human genome fragm
5	38.6	15.0	6354	19 V69699	5' UTR sequence of
6	38.2	14.8	4228	21 Z51683	Human cyclic nucle
7	37.6	14.6	170	19 V69706	Nucleotide sequenc
8	36.4	14.1	1526	20 Z42235	Human normal bladd
9	36.2	14.0	1000	21 A02484	Human colon cancer
10	34.8	13.5	2953	19 V62462	Rat transcription
11	34.2	13.3	3690	13 Q30849	Type III procollag
12	33.8	13.1	4539	21 A08372	Human L-type amino

C 13	33.6	13.0	1120	15	Q70728
C 14	33.6	13.0	1120	17	T42214
C 15	33.6	13.0	1120	18	T79601
C 16	33.6	13.0	3765	13	Q31890
C 17	33.6	13.0	5460	17	T16508
C 18	33	12.8	1072	13	Q24285
C 19	33	12.8	11722	19	V34455
C 20	33	12.8	114955	20	X53491
C 21	32.8	12.7	513	20	X91237
C 22	32.8	12.7	514	20	X91400
C 23	32.8	12.7	1450	21	Z47131
C 24	32.8	12.7	8169	19	V26609
C 25	32.4	12.6	201	8	N70194
C 26	32.4	12.6	201	8	N70195
C 27	32.4	12.6	2681	20	X58429
C 28	32.2	12.5	1524	21	A08393
C 29	32	12.4	742	21	Z45675
C 30	32	12.4	1459	21	A02528
C 31	32	12.4	3410	19	V61201
C 32	32	12.4	3410	19	V58586
C 33	32	12.4	3410	21	A06349
C 34	32	12.4	34094	20	Z30463
C 35	31.8	12.3	599	19	V43880
C 36	31.8	12.3	1321	21	A37128
C 37	31.6	12.2	1586	19	V34287
C 38	31.6	12.2	1907	19	V34314
C 39	31.6	12.2	2074	19	V32932
C 40	31.6	12.2	2074	19	V32933
C 41	31.4	12.2	595	14	Q50041
C 42	31.4	12.2	1335	13	Q23295
C 43	31.4	12.2	13842	21	Z87297
C 44	31.4	12.2	36778	21	Z87318
C 45	31.4	12.2	37948	21	Z87285

ALIGNMENTS

RESULT 1
T94471
ID T94471 standard; cDNA; 1228 BP.
AC T94471;
DT 03-MAR-1998 (first entry)
XX Human Fchd605 gene differentially regulated in monocytes.
DE Fchd602 gene; differential expression; monocyte; human;
KW foam cell; cardiovascular disease; atherosclerosis; ~~th~~haemia;
KW reperfusion; hypertension; restenosis; arterial inflammation;
KW therapy; diagnosis; drug screening; marker; ss.
XX Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..471
FT tag= a

TaTA-binding prote
Drosophila TATA-bi
TATA-binding prote
FMR-1 gene for det
Vector PAC3A1 cont
Epo:IL-3 Flex, rec
Human MHC class I
Human adenosine A1
T. gondii immunoge
T. gondii immunoge
Pseudomonas aerugi
Actinomodura hibis
Signal portion of
Streptomyces prote
Thermophilus therm
Human L-type amino
cDNA sequence of a
Human colon cancer
Full length cDNA s
Prostate tumour sp
Human immunogenic
Complete nucleotid
Mycobacterial meth
Human PRO1781 (UNQ
Human secreted pro
Human secreted pro
Human cyclin-depen
Human cyclin-depen
ICP34.5 fragment.
HSV-1 (F) ICP34.5
S. venezuelae macr
S. venezuelae pik
S. venezuelae pik

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 10:53:09 ; Search time 1450.12 Seconds
(without alignments)
5934.126 Million cell updates/sec

Title: US-08-799-910-9
Perfect score: 1228
Sequence: 1 ATCTGTCACCTCCGAGCTG.....AAAAAAAAAACTCGAG 1228

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*
48: gb_est48.*
49: gb_est49.*
50: gb_est50.*
51: gb_est51.*
52: gb_est52.*
53: gb_est53.*
54: gb_est54.*
55: gb_est55.*
56: gb_est56.*
57: gb_est57.*
58: gb_est58.*
59: gb_est59.*
60: gb_est60.*
61: gb_est61.*
62: gb_est62.*
63: gb_est63.*
64: gb_est64.*
65: gb_est65.*
66: gb_est66.*
67: gb_est67.*
68: gb_est68.*
69: gb_est69.*
70: gb_est70.*
71: gb_est71.*
72: gb_est72.*
73: gb_est73.*
74: gb_est74.*
75: gb_est75.*
76: gb_est76.*
77: gb_est77.*
78: gb_est78.*
79: gb_est79.*
80: gb_est80.*
81: gb_est81.*
82: gb_est82.*
83: gb_est83.*
84: gb_est84.*
85: gb_est85.*
86: gb_est86.*
87: gb_est87.*
88: gb_est88.*
89: gb_est89.*
90: gb_est90.*
91: gb_est91.*
92: gb_est92.*
93: gb_est93.*
94: gb_est94.*
95: gb_est95.*
96: gb_est96.*
97: gb_est97.*
98: gb_est98.*
99: gb_est99.*
100: gb_est100.*
101: gb_est101.*
102: gb_est102.*
103: gb_est103.*
104: gb_est104.*
105: gb_est105.*
106: gb_est106.*
107: gb_est107.*
108: gb_est108.*
109: gb_est109.*
110: gb_est110.*
111: gb_est111.*
112: gb_est112.*
113: gb_est113.*
114: gb_est114.*
115: gb_est115.*
116: gb_est116.*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estpl9:*
121: em_estpl10:*
122: em_estpl11:*
123: em_estpl12:*
124: em_estpl13:*
125: em_estpl14:*
126: em_estpl15:*
127: em_estpl16:*
128: em_estpl17:*
129: em_estpl18:*
130: em_estpl19:*
131: em_estpl20:*
132: em_estpl21:*
133: em_estpl22:*
134: em_estpl23:*
135: em_estpl24:*
136: em_estpl25:*
137: em_estpl26:*
138: em_estpl27:*
139: em_estpl28:*
140: em_estpl29:*
141: em_estpl30:*
142: em_estpl31:*
143: em_estpl32:*
144: em_estpl33:*
145: em_estpl34:*
146: em_estpl35:*
147: em_estpl36:*
148: em_estpl37:*
149: em_estpl38:*
150: em_estpl39:*
151: em_estpl40:*
152: em_estpl41:*
153: em_estpl42:*
154: em_estpl43:*
155: em_estpl44:*
156: em_estpl45:*
157: em_estpl46:*
158: em_estpl47:*
159: em_estpl48:*
160: em_estpl49:*
161: em_estpl50:*
162: em_estpl51:*
163: em_estpl52:*
164: em_estpl53:*
165: em_estpl54:*
166: em_estpl55:*
167: em_estpl56:*
168: em_estpl57:*
169: em_estpl58:*
170: em_estpl59:*
171: em_estpl60:*
172: em_estpl61:*
173: em_estpl62:*
174: em_estpl63:*
175: em_estpl64:*
176: em_estpl65:*
177: em_estpl66:*
178: em_estpl67:*
179: em_estpl68:*
180: em_estpl69:*
181: em_estpl70:*
182: em_estpl71:*
183: em_estpl72:*
184: em_estpl73:*
185: em_estpl74:*
186: em_estpl75:*
187: em_estpl76:*
188: em_estpl77:*
189: em_estpl78:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	748.2	60.9	951	107	BE383865	BE383865 601273192
2	621.4	50.6	765	106	BE275966	BE275966 601120727
3	604.2	49.2	762	110	BE621346	BE621346 601493736
4	590	48.0	916	137	BE882130	BE882130 601505227
5	589.4	48.0	603	106	BE273386	BE273386 601143493
6	588	47.9	810	109	BE563592	BE563592 601334706
7	585	47.6	585	137	BE910077	BE910077 601502143
8	581.2	47.3	586	89	AW340002	AW340002 hc91d03.x
9	579.6	47.2	598	93	AW662348	AW662348 hi25c07.x
10	567.2	46.2	614	110	BE621600	BE621600 601493736
11	561.6	45.7	588	97	AW956285	AW956285 EST368355
12	557	45.4	617	38	AW690660	AW690660 AV690660
13	544.8	44.4	548	88	AW263252	AW263252 xn79q02.x
14	540.8	44.0	1070	137	BE878459	BE878459 601487866
15	519.4	42.3	521	25	A1800828	A1800828 wg13f12.x
16	517.4	42.1	544	107	BE384949	BE384949 601276496
17	515.6	42.0	523	18	A1302813	A1302813 qn58g09.x
18	514.8	41.9	585	97	AW960290	AW960290 EST372361
19	513.4	41.8	518	12	AA806234	AA806234 oe29h02.s
20	512	41.7	513	40	AW150706	AW150706 xg38a10.x
21	506.2	41.2	524	22	A1566501	A1566501 tr60f07.x
22	505.4	41.2	509	9	AA582942	AA582942 nn80c01.s
23	503.4	41.0	521	23	A1685453	A1685453 tr88d12.x
24	501.6	40.8	510	89	AW337920	AW337920 xw63a07.x
25	496.4	40.4	553	142	N32077	N32077 yw97d08.s1
26	494.8	40.3	1155	135	BE786915	BE786915 601477783
27	487.6	39.7	548	96	AW887044	AW887044 RC1-OT008
28	486.8	39.6	535	25	A1802925	A1802925 tj58a07.x
29	486.6	39.6	494	15	A1022951	A1022951 ow66g09.s
30	482.4	39.3	484	19	A1336188	A1336188 ql44c12.x
31	481.4	39.2	491	39	AW081284	AW081284 xc30c01.x
32	481	39.2	492	10	AA643850	AA643850 np26f08.s
33	478.4	39.0	480	96	AW874420	AW874420 hq04a02.x
34	472.4	38.5	496	16	A1160053	A1160053 qc08a02.x
35	472	38.4	475	8	AA512938	AA512938 nh91b09.s
36	470.8	38.3	487	23	A1687081	A1687081 tp92h08.x
37	467	38.0	701	135	BE735733	BE735733 601304556
38	464.6	37.8	483	16	A1091958	A1091958 qa59d10.s
39	461.4	37.6	466	39	AW051587	AW051587 wv87h05.x
40	458.4	37.3	460	17	A1185199	A1185199 qc35h09.s
41	458.4	37.3	460	24	A1738521	A1738521 w133a06.x
42	458.4	37.3	463	13	AA884985	AA884985 am35e07.s
43	453	36.9	461	12	AA812286	AA812286 nr82f07.s
44	451.6	36.8	459	147	W60982	W60982 zc98h09.s1
45	450	36.6	469	24	A1761431	A1761431 wq65e08.x

ALIGNMENTS

RESULT 1
BE383865
LOCUS 601273192F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3614255 5',
DEFINITION BE383865 951 bp mRNA ESI
ACCESSTON BE383865
VERSION BE383865.1 GI:9329230
KEYWORDS EST.
SOURCE human.


```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 951)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC/DC/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM276 row: 9 column: 24
High quality sequence stop: 747.
FEATURES
    source
        1..951
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3614235"
            /clone_lib="NIH-MGC_20"
            /tissue_type="melanotic melanoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dr priming. Directionally
            adaptor: GGCACGAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT
ORIGIN

```
Query Match 60.9%; Score 748.2; DB 107; Length 951;
Best Local Similarity 91.5%; Pred. No. 4.8e-180;
Matches 848; Conservative 0; Mismatches 73; Indels 6; Gaps 5;

Qy 2 TGTGTCTACCTCGCAGCTGCCACCGACCATCACCATCTCGCAGCGCCCGACCGCCGCC 61
Db 1 TGTGTCTACCTCGCAGCTGCCACCGACCATCACCATCTCGCAGCGCCCGACCGCCGCC 60
Qy 62 CTCTCACCATTCCGGGACCGCGGGGCTCGGTCTCTGAGATCTTACCTTCGACCCCTC 121
Db 61 CTCTCACCATTCCGGGACCGCGGGGCTCGGTCTCTGAGATCTTACCTTCGACCCCTC 120
Qy 122 TCCCGGAGCCCGCAGCGGCCCTCGCGGCGCCCGCAGCGCCCTCTCGCGGCGACCGAAAC 181
Db 121 TCCCGGAGCCCGCAGCGGCCCTCGCGGCGCCCGCAGCGCCCTCTCGCGGCGACCGAAAC 180
Qy 182 GCAGCCCGCCAGGGTCTCTACCTCGAGTGTCTCGCGCGCCAGCTGCCAGTCGAGGAACCGA 241
Db 181 GCAGCCCGCCAGGGTCTCTACCTCGAGTGTCTCGCGCGCCAGCTGCCAGTCGAGGAACCGA 240
Qy 242 ACCCAGCCAAAGGCTTCTCTTCTGCTCTCACCATCTCTCTGCGCAGATCTTGATGG 301
Db 241 ACCCAGCCAAAGGCTTCTCTTCTGCTCTCACCATCTCTCTGCGCAGATCTTGATGG 300
Qy 302 CTGAAGAGGTGTGCGCGGGGCTCTGCTCTCGAGAGCGCCCTTACGCGCGATCCCTCGG 361
Db 301 CTGAAGAGGTGTGCGCGGGGCTCTGCTCTCGAGAGCGCCCTTACGCGCGATCCCTCGG 360
Qy 362 CGCCCGCCCGCTGTCTCCCGCTCTCGAGCCCTTTTAACTGACTTCGGAGCCCTCGGACT 421
Db 361 CGCCCGCCCGCTGTCTCCCGCTCTCGAGCCCTTTTAACTGACTTCGGAGCCCTCGGACT 420
Qy 422 ACAGCTCTGAGACCTCAGACCTTCTCTCAGACACCGCGCGCTTCTTACTGTGACTCC 481
Db 421 ACAGCTCTGAGACCTCAGACCTTCTCTCAGACACCGCGCGCTTCTTACTGTGACTCC 480
```

```
Qy 482 CCGCACTCCCAAAAGAAATCCGAAACACCAAGAAACACCAAGAGGCTACCTGTCGCG 541
Db 481 CCGCACTCCCAAAAGAAATCCGAAACACCAAGAAACACCAAGAGGCTACCTGTCGCG 540
Qy 542 GAGAGCGTATCCCAACTGGGACTTCCGAGCACTTGAATCACTACACACTACAGCCGAG 601
Db 541 GAGAGCGTATCCCAACTGGGACTTCCGAGCACTTCCGAGCACTTGAATCACTACAGCCGAG 599
Qy 602 ACGCCACCGCGTCTTGAGCGCGGACCGAGCGGCACAGAGACCGAGCGGCATAGAGACCG 661
Db 600 ACGCCACCGCGTCTTGAGCGCGGACCGAGCGGCACAGAGACCGAGCGGCATAGAGACCG 659
Qy 662 AGGCACACCGCCAGC-TGGGGCTAGCGCCCGTGGGAAGAGAGAGCGTCTTAACTTATCT 720
Db 660 AGGCACACCGCCAGC-TGGGGCTAGCGCCCGTGGGAAGAGAGAGCGTCTTAACTTATCT 719
Qy 721 TATTGCTCTTAATTAATTAATTAATTAATG--TATTTATGACGTCCTTCTAGGTAGGAGAT 778
Db 720 TATAGCTCTTAATTAATTAATTAATTAATTAATG--TATTTATGACGTCCTTCTAGGTAGGAGAT 779
Qy 779 GTGTACGTAAATTAATTAATTAATTAATTAATGAGGTGTGAGATGTTCCCTCTGCTGTAAT 838
Db 780 GTGTCCGTCACAAATTA-TCAAACCTTATGCA--GGTGTGACATGTTTCCCTCTGTTGAAAGC 837
Qy 839 GCAGTCTCTTGGTATTTATTGAGCTTTGTGGGACTGCTGGAAGCAGACACCTGCAACT 898
Db 838 AGGCACAGGATTAATGAGCTTGGGACTGCTGGAAGCAGACACCTGGAAGACT 897
Qy 899 CGCGCAAAAGTAGGAGAGAAATGGGGA 925
Db 898 AGGGAAGACCGCGGACCAAGATGGGGA 924

RESULT 2
BE275966 765 bp mRNA 13-JUL-2000
LOCUS 601120727P1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2967534 5',
DEFINITION mRNA sequence.
ACCESSION BE275966
VERSION BE275966.1 GI:9150928
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 765)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC/DC/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM272 row: e column: 07
High quality sequence start: 7
High quality sequence stop: 699.
Location/Qualifiers
    1..765
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:2967534"
        /clone_lib="NIH_MGC_20"
        /tissue_type="melanotic melanoma"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
        EcoRI; cDNA made by oligo-dr priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
```


QY 598 GGAGACCCACCGGTCTTTCAG-GCGGACCGA-GGCCACACAGACCCGAGC-CCATA 654
 DB 600 GAGCACCCACCGGTCTTTCAGCGCGGACCGACCGGACAGACAGAGCGGCATA 659
 QY 655 GAGACCGAGCAGACCCAGCTGGGCTAGGCGCGGTGGAGAGAGAGCGTCTTAATTT 714
 DB 660 GAGACCGAGCAGACAGACAGCTGGGCTAGGCGCGGTGGAGAGAGAGCGTCCGCAAT 719
 QY 715 ATTCTTATTGCTCTTAATTAATTAATATAT 745
 DB 720 ATTACTTATGCTCTTAATTAATTAATAGGTAT 750

RESULT 4
 BE882130
 LOCUS 916 bp mRNA EST 27-SEP-2000
 DEFINITION 601505227F2 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906823 5',
 mRNA sequence.
 ACCESSION BE882130
 VERSION BE882130
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM703 row: f column: 08
 High quality sequence stop: 659.

FEATURES
 source
 1..916
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3906823"
 /clone_lib="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SpOUT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."
 BASE COUNT 187 a 327 c 268 g 134 t
 ORIGIN

Query Match 48.0%; Score 590; DB 137; Length 916;
 Best Local Similarity 94.7%; Pred. No. 9.4e-140;
 Matches 676; Conservative 0; Mismatches 30; Indels 8; Gaps 6:

QY 31 ATGACCATCTTCAGCGCCGACCGCGCGCCCTCCACCATCCGGGACCCCGCGGGGC 90
 DB 1 ATGACCATCTTCAGCGCCGACCGCGCGCCCTCCACCATCCGGGACCCCGCGGGGC 59
 QY 91 TCGGTCTTCAGATCTTACCTTCGACCTCTCCGGAGCCCGACGCGCCCTCGCGG 150
 DB 60 TCGGTCTTCAGATCTTACCTTCGACCTCTCCGGAGCCCGACGCGCCCTCGCGG 119
 QY 151 CGCCCGACGCGCTCTTCGCGGGACCGAAGACGAGCGCGAGGTTCTTACCTCGAGTG 210
 DB 120 CGCCCGACGCGCTCTTCGCGGGACCGAAGACGAGCGCGAGGTTCTTACCTCGAGTG 179
 QY 211 GTCCGGCGCAGCTGCCACTCGAGGAGACCGAACCCAGCAGCAAGAGCTTCCTCTCTGCTG 270

DB 180 GTCCGGCGCAGCTGCCAGTCGAGGAACCGAACCCAGCCAAAGAGCTTCCTTCCTGCTG 239
 QY 271 CTCACCATCTCTTCTGCGCAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCGCCCTGCCT 330
 DB 240 CTCACCATCTCTTCTGCGCAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCGCCCTGCCT 299
 QY 331 CCAGGAGCGCCCTTAACCGCCGATCCCTGGGCGCCACCCCTGTGTCTCCCTCTCTCGAG 390
 DB 300 CCAGGAGCGCCCTTAACCGCCGATCCCTGGGCGCCACCCCTGTGTCTCCCTCTCTCGAG 359
 QY 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGCTCAGACTTCTCTCCAG 450
 DB 360 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGCTCAGACTTCTCTCCAG 419
 QY 451 CAACACCGCGC-CGCTTCTTAATCTGACTTCCTCCGACTCCCAAAAAGAAATCCGAATA 509
 DB 420 CAACACCGCGCAGCGCTTCTTAATCTGACTTCCTCCGACTCCCAAAAAGAAATCCGAATA 479
 QY 510 CCACAAGAAACACACGAGCGGTACTCTGGCGGAGAGCTATCCCAACTGGGACTTCGCG 569
 DB 480 CCACAAGAAACACGAGCGGTACTCTGGCGGAGAGCTATCCCAACTGGGACTTCGCG 539
 QY 570 AGCCAACTTGAATCTCAGAACTACAGCGGAGACCCAGCCCGTGTCTGAGCGGGACCG 629
 DB 540 AGCCAACTTGAATCTCAGAACTACAGCGGAGACCCAGCCCGTGTCTGAGCGGGACCG 597
 QY 630 AGCGCACAGACCGCGCGCTAGAGACCGAGGCGACAGCCAGCTGGGCTAGGCGCG 689
 DB 598 AGCGCACAGACCGCGCGCTAGAGACCGAGCGCATAGAGA-CGAGGCACAGCC--AGCTGGCGCTAGCGCG 653
 QY 690 GTGGGAAGGAGAGCGCTGCTTAATTTATTTCTTATTTCTGCTTAATTAATTTAT 743
 DB 654 GTGGGAAGGAGAGCGCTGCTTAATTTATTTCTTATTTCTGCTTAATTAATTTAT 707

RESULT 5

BE273386
 LOCUS 603 bp mRNA EST 13-JUL-2000
 DEFINITION 601143493F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:3507191 5',
 mRNA sequence.
 ACCESSION BE273386
 VERSION BE273386
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: DCTD/DBP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM186 row: j column: 24
 High quality sequence stop: 579.

FEATURES

source

1..603
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3507191"
 /clone_lib="NIH_MGC_14"
 /tissue_type="renal cell adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: kidney; Vector: pONH7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally


```

QY 540 GCGAGAGCGTATCCCACTGG-GACTCTCGAGGCAACTTGAACCTCAGAAC-ACTACAGC 597
|||||
DB 535 GCGAGAGCGTATCCCACTGGAGACTTCGAGGCAACTTGAACCTCAGAACACTACAGC 594
|||||
QY 598 GGAGAGCGCCACACCC-GGTGCTTGAGCGGAGCGAGCGGCA-CAGAGACCGAGG-CGCAT 653
|||||
DB 595 GGAGAGCGCCACACCTGGTGGTCTTGAGCGGAGCGGAGCGG77-CAAGAGACCGAGCGCGCAT 654
|||||
QY 654 AGAGACCGAGGCAACCCAGCTGGGGCTAGGCCCGTGGGAGGAGGACCGCTCGTTAAAT 713
|||||
DB 655 AGAGACCGAGGCAACAG-CCAGATGGGCGCTAGGCCCGTGGGAGGAGAGCGTCCGCAAT 713
|||||
QY 714 TATTTCTTATGCTCTCTAATTAATATATATATATATATATATATATATATATATATATATAT 773
|||||
DB 714 TATATATCTTATGCTCTCTAATTAATATATATATATATATATATATATATATATATATATATAT 772
|||||
QY 774 GAGATGCTGACGTAATATATATATATATATATATATATATATATATATATATATATATATAT 811
|||||
DB 773 GACAAATGTCAGTAATATATATATATATATATATATATATATATATATATATATATATATAT 810
|||||

RESULT 7
BE910077 585 bp mRNA 29-SEP-2000
LOCUS 601502143F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3903754 5',
DEFINITION mRNA sequence.
ACCESSION BE910077
VERSION BE910077.1 GI:10406309
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 585)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM695 row: f column: 11
High quality sequence stop: 585.
FEATURES
source
1..585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3903754"
/lab_host="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 131 a 119 c 188 g 147 t
ORIGIN
Query Match 47.6%; Score 585; DB 137; Length 585;
Best Local Similarity 100.0%; Pred. No. 1.7e-138;
Matches 585; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 544 GACCGTATCCCACTGGGACTTCGAGGCAACTTGAACCTCAGAACACTACAGCGAGAC 603
|||||
DB 1 GAGCGTATCCCACTGGGACTTCGAGGCAACTTGAACCTCAGAACACTACAGCGAGAC 60
|||||

```

```

QY 604 GCCACCCGCTGTAGCGCGGACCGACAGAGACCGAGGCGCATAGAGACCGAG 663
|||||
DB 61 GCCACCCGCTGTAGCGCGGACCGCAAGCGCACAGAGACCGAGCGCATAGAGACCGAG 120
|||||
QY 664 GCACAGCCAGCTGGGGCTAGGCCCGTGGGAGGAGAGGCTGTTAATTTATTTCTTAT 723
|||||
DB 121 GCACAGCCAGCTGGGGCTAGGCCCGTGGGAGGAGAGGCTGTTAATTTATTTCTTAT 180
|||||
QY 724 TCTCTCTAAATTAATATTTATATATATATATATATATATATATATATATATATATAT 783
|||||
DB 181 TCTCTCTAAATTAATATTTATATATATATATATATATATATATATATATATATATAT 240
|||||
QY 784 CGTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 843
|||||
DB 241 CGTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 300
|||||
QY 844 TCTCTCTGGTATTTATTTAGCTTTGTGGGACTGTGTGAAGCAGGACACCTTGGAACTCGGC 903
|||||
DB 301 TCTCTCTGGTATTTATTTAGCTTTGTGGGACTGTGTGAAGCAGGACACCTTGGAACTCGGC 360
|||||
QY 904 AAGTAGGAGAGAAATGGGAGGACTCGGCTGGGGAGGAGCGTCCCGGCTGGGATCAAG 963
|||||
DB 361 AAGTAGGAGAGAAATGGGAGGACTCGGCTGGGGAGGAGCGTCCCGGCTGGGATCAAG 420
|||||
QY 964 TCTGTGTGGTGGTCTGAAGTTTAGGAGGCTGACTGCATCTCCAGCATCTCAACTCCGCT 1023
|||||
DB 421 TCTGTGTGGTGGTCTGAAGTTTAGGAGGCTGACTGCATCTCCAGCATCTCAACTCCGCT 480
|||||
QY 1024 GTCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCCTTCCATCT 1083
|||||
DB 481 GTCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCCTTCCATCT 540
|||||
QY 1084 TCTTGAAGTCGCTTTAGGCTGGCTCGGAGTAGAGGTTGGGG 1128
|||||
DB 541 TCTTGAAGTCGCTTTAGGCTGGCTCGGAGTAGAGGTTGGGG 585
|||||

RESULT 8
AW340002/c 586 bp mRNA EST 31-JAN-2000
LOCUS hc91d03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2907365 3', mRNA sequence.
ACCESSION AW340002
VERSION AW340002.1 GI:6836628
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 586)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 471.
FEATURES
source
1..586
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2907365"
/lab_host="Soares_NFL_T_GBC_S1"
/Note="Organ: pooled; Vector: p773D-pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NIH, and B-cell

```

NCI-CCAP_GCB1} were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 155 a 187 c 110 g 134 t

Query Match 47.3%; Score 581.2; DB 89; Length 586;
Best Local Similarity 99.5%; Pred. No. 1.5e-137;
Matches 583; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 621 GCGGGACCGGCGGCACAGACCGGCGCATAGAGCCGAGGACAGCCAGCTGGGG 680
DB 586 GCGGGACCGGCGGCACAGACCGGCGCATAGAGCCGAGGACAGCCAGCTGGGG 527
QY 681 CTAGGCGCGGTGGGAAGGAGCGTCTTAATTTATTTCTTATTTCTCTTAATTAAT 740
DB 526 CTAGGCGCGGTGGGAAGGAGCGTCTTAATTTATTTCTTATTTCTCTTAATTAAT 467
QY 741 TATATCTATTTATGTACGCTCTCTAGGTGATGCGAGATGTACGTAATATTTTAA 800
DB 466 TATATCTATTTATGTACGCTCTCTAGGTGATGCGAGATGTACGTAATATTTTAA 407
QY 801 CTATCCAGGGTGTGAGATGTCTCTGCTGTAATGAGCTCTCTTGGTATTATG 860
DB 406 CTATCCAGGGTGTGAGATGTCTCTGCTGTAATGAGCTCTCTTGGTATTATG 347
QY 861 AGCTTTGTGGGACTGTGGAAGCAGCAGACCTGGACTCGGCAAGTAGGAGAGAAAT 920
DB 346 AGCTTTGTGGGACTGTGGAAGCAGCAGACCTGGACTCGGCAAGTAGGAGAGAAAT 287
QY 921 GGGGAGGACTCGGGTGGGGAGGAGCTCCCGGCTGGGAGTGAAGTCTGCGTGGGCGTA 980
DB 286 GGGGAGGACTCGGGTGGGGAGGAGCTCCCGGCTGGGAGTGAAGTCTGCGTGGGCGTA 227
QY 981 AGTTTAGGAGGACTCGGATCTCCAGCATCTCAACTCCGCTGCTGCTACTGTGAGAT 1040
DB 226 AGTTTAGGAGGACTCGGATCTCCAGCATCTCAACTCCGCTGCTGCTACTGTGAGAT 167
QY 1041 TCGGCGGACATTTAGGAATGAGATCCGTGAGATCTTCCATCTCTTGAAGTGGCTTTA 1100
DB 166 TCGGCGGACATTTAGGAATGAGATCCGTGAGATCTTCCATCTCTTGAAGTGGCTTTA 107
QY 1101 GGGTGGCTGCGAGGTAGAGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGT 1160
DB 106 GGGTGGCTGCGAGGTAGAGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGT 47
QY 1161 GCGCTAGTATGTTCTGTGAACACAAATAAATTTGATTTACTGTCAA 1206
DB 46 GCGCTAGTATGTTCTGTGAACACAAATAAATTTGATTTACTGTCAA 1

RESULT 9
LOCUS AW662348/c 598 bp mRNA EST 06-APR-2000
DEFINITION hi25c07.x1 NCI-CCAP_Col4 Homo sapiens cDNA clone IMAGE:2973324 3',
mRNA sequence.
ACCESSION AW662348
VERSION AW662348.1 GI:7454886
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 598)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CCAP),
TITLE Tumor Gene Index

JOURNAL COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA library Preparation: Life Technologies, Inc.

DNA library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

Seq primer: -400P from Gibco

High quality sequence stop: 392.

Location/Qualifiers

SOURCE

1. 598
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2973324"
/clone_lib="NCI-CCAP_Col4"
/issue_type="moderately-differentially-adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT6; Sile_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

BASE COUNT 163 a 187 c 116 g 132 t

ORIGIN

Query Match 47.2%; Score 579.6; DB 93; Length 598;
Best Local Similarity 98.5%; Pred. No. 3.9e-137;
Matches 585; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 611 GGTGCTTGAGCGGACCGGCGGCACAGACCGGCGCATAGAGCCGAGGACAGCCAGCAGC 670
DB 598 GGTGCTTGAGCGGACCGGCGGCACAGACCGGCGCATAGAGCCGAGGACAGCCAGCAGC 539
QY 671 CCAGCTGGGGTAGGCGGCGGGAAGGAGGAGCGTCTTAATTTATTTCTTATTTCTCTCT 730
DB 538 CCAGCTGGGGTAGGCGGCGGGAAGGAGGAGCGTCTTAATTTATTTCTTATTTCTCTCT 479
QY 731 AATTAATATTTATATGATTTATGACCTCTCTCTAGTGTAGTGTAGTGTAGTGTAGTGTAG 790
DB 478 AATTAATATTTATATGATTTATGACCTCTCTCTAGTGTAGTGTAGTGTAGTGTAGTGTAG 419
QY 791 TTTATTTAACTTTATGCAAGGCTGTGAGATGTCTCTCTCTGTAATGCAAGTGTCTCTTG 850
DB 418 TTTATTTAACTTTATGCAAGGCTGTGAGATGTCTCTCTCTGTAATGCAAGTGTCTCTTG 359
QY 851 GTATTTATGAGCTTTGTGGAGCTGGTGAAGGAGGAGACCTGGAACTGGGCAAAATAG 910
DB 358 GTATTTATGAGCTTTGTGGAGCTGGTGAAGGAGGAGACCTGGAACTGGGCAAAATAG 299
QY 911 GAGAAGAAATCGGAGGACTCGGGTGGGGGAGGAGCTCCCGGCTCCGATGAGTCTCTTG 970
DB 298 GAGAAGAAATCGGAGGACTCGGGTGGGGGAGGAGCTCCCGGCTCCGATGAGTCTCTTG 239
QY 971 GTGGGTCTTAAGTTTAGGAGTGTGATCTCTCCAGCATCTCAACTCCGCTCTGCTACT 1030
DB 238 GTGGGTCTTAAGTTTAGGAGTGTGATCTCTCCAGCATCTCAACTCCGCTCTGCTACT 179
QY 1031 GTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCTCTTCCATCTTCTTGA 1090
DB 178 GTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCTCTTCCATCTTCTTGA 119
QY 1091 GTCCGCTTTAGGCTGGCTGGCAGGTAGAGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGG 1150
DB 118 GTCCGCTTTAGGCTGGCTGGCAGGTAGAGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGG 59
QY 1151 CTGTCGAGATCGCCTAGTATGTTCTGTGAACACAAATAAATTTGATTTACTGTGTC 1204
DB 58 CTGTCGAGATCGCCTAGTATGTTCTGTGAACACAAATAAATTTGATTTACTGTGTC 5

```

RESULT 10
BE621600/c
LOCUS
DEFINITION
HE621600 614 bp mRNA EST 24-AUG-2000
601493736T1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3895736 3',
mRNA sequence.
BE621600
VERSION
BE621600.1 GI:9892540
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 614)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue procurement: ATCC
CDNA Library preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM674 row: h column: 09
High quality sequence start: 12
High quality sequence stop: 613.
High quality sequence stop: 613.

FEATURES
source
1..614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3895736"
/clone_lib="NIH_MGC_70"
/tissue_type="epitheloid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 158 a 198 c 123 g 135 t
ORIGIN

Query Match 46.2%; Score 567.2; DB 110; Length 614;
Best Local Similarity 98.2%; Pred. No. 5.7e-134;
Matches 595; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

QY 582 CTCAGAACACTACAGCGGAGACGCCACCCGGTGTCTTGAGCGGCGACCGGCACAGAG 641
|||||
DB 614 CTCAGAACACTACAGCGGAGACGCCACCCGGTGTCTTGAGCGGCGACCGGCACAGAG 555
|||||
QY 642 ACCGAGCGCATAGACCGAGCGACAGCCAGCTGGGCTAGGCCGTGGGGAAGAGA 701
|||||
DB 554 ACCGAGCGCATAGACCGAGCGACAGCCAGCTGGGCTAGGCCGTGGGGAAGAGA 495
|||||
QY 702 GCGTCGTAA--TTTATTCTTATTGCTCTCTAATTAATTTATATATTTATGTACGT 759
|||||
DB 494 GCGTCGTAA--TTTATTCTTATTGCTCTCTAATTAATTTATATATTTATGTACGT 435
|||||
QY 760 CTTCTTACGTGATGACATGTGTACATAATTTATTTTAACTTATGCAAGGTGTGAGA 819
|||||
DB 434 CTTCTTACGTGATGACATGTGTACATAATTTATTTTAACTTATGCAAGGTGTGAGA 375
|||||
QY 820 TGTTCCTCTGCTGTAATGTCAGGTCTCTTGTGATTTATTTAGCTTTGTGGGACTGGTG 879
|||||
DB 374 TGTTCCTCTGCTGTAATGTCAGGTCTCTTGTGATTTATTTAGCTTTGTGGGACTGGTG 315
|||||
QY 880 AAGCAGGACACCTGGGAACCTGCGCAAGTAGGAGAGAATGGGAGGA-CTCGGCTGGG 938
|||||

```

```

DB 314 AAGCAGGACACCTGGAACTCGGCAAAAGTAGGAGAGAAATGGGAGGATCTCGGGTGGG 255
QY 939 GGAGGACGTCCTCCGCTGGGATGAAGTCTGCTGGTGGTCTCTAAGTCTAGGAGGTGACTCC 998
|||||
DB 254 GGAGGACGTCCTCCGCTGGGATGAAGTCTGCTGGTGGTCTCTAAGTCTAGGAGGTGACTCC 195
|||||
QY 999 ATCTCTCAGCATCTCAACTCCGCTGTCTACTGTGTGAGACTTCGGCGGACCATTAGGAA 1058
|||||
DB 194 ATCTCTCAGCATCTCAACTCCGCTGTCTACTGTGTGAGACTTCGGCGGACCATTAGGAA 135
|||||
QY 1059 TGAGATCCGTGAGATCCTTCCATCTTCTTGAAGTCGCCCTTTAGGGTGGCGGAGGTAGA 1118
|||||
DB 134 TGAGATCCGTGAGATCCTTCCATCTTCTTGAAGTCGCCCTTTAGGGTGGCGGAGGTAGA 75
|||||
QY 1119 GGGTTGGGGTGTGGTGGCTGTCCAGGAGCGACTTCGAGATCGCCTACTATCTTCTGTG 1178
|||||
DB 74 GGGTTGGGGTGTGGTGGCTGTCCAGGAGCGACTTCGAGATCGCCTACTATCTTCTGTGGA 15
|||||
QY 1179 AACACA 1184
|||||
DB 14 CACAAA 9

RESULT 11
AW956285
LOCUS
DEFINITION
AW956285 588 bp mRNA EST 01-JUN-2000
EST368355 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
ACCESSION
AW956285
VERSION
AW956285.1 GI:8145968
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 588)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holl,
J. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 90
Seq primer: Reverse.
Location/Qualifiers
1..588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGD"
/notes="Vector: pBluescriptSKm"
BASE COUNT 135 a 117 c 187 g 149 t
ORIGIN

Query Match 45.7%; Score 561.6; DB 97; Length 588;
Best Local Similarity 98.3%; Pred. No. 1.5e-132;
Matches 578; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 556 AACTGGGACTTCGGAGGCAACTTGAACTCAGAACACTACAGCGGAGACGCCACCCCGTGC 615
|||||
DB 1 AACTGGGACTTCGGAGGCAACTTGAACTCAGAACACTACAGCGGAGACGCCACCCCGTGC 60
|||||
QY 616 TTGAGCGCGGACCGGCGCACAGAGACCGGCGCATAGAGACCGGAGGACAGCCAGCCAGC 675
|||||
DB 61 TTGAGCGCGGACCGGCGCACAGAGACCGGCGCATAGAGACCGGAGGACAGCCAGCCAGC 120
|||||
QY 675 TGGGGCTAGGCGCGGAGGAGCGTCTGTTATTTATTTATTTATTTATTTATTTATTTATTT 735
|||||

```


Ddb	121	"TGGCC"TAGCCCCGTCGGGAAGAGAGCGTCGTAA"TA"TTATTCTTATTGC"TCCTAATTA	180
Oy	736	ATATTTATATGTATTTATGTACGTCTCCTAGGTGATGGAGATGCTGACATAATATTTAT	795
Ddb	181	ATATTTATATGTATTTATGTACGTCTCCTAGGTGATGGAGATGCTGACATAATATTTAT	240
Oy	796	"TTTAACT"ATGCAAGCGTGTGAGATGTTCCTCTCTGCTGTAAT"GCAGGTCCTCTTTCGTATT	855
Ddb	241	TTTAACTTTATGCAAGGGTGTGAGATGTTCCCTCTGCTGTAAATGCAGGTCCTCTTGGTATT	300
Oy	856	TATTGACCTTTGTGGACTGTGTGAAGCAGGACACCTTGGAACTCGGCAAGTAGAGAA	915
Ddb	301	TATTGACCTTTGTGGACTGTGTGAAGCAGGACACCTTGGAACTCGGCAAGTAGAGAA	360
Oy	916	GAATCGGAGAGACTCGGGTGGGGAGGACGTCCGGGTGGGATGAAGTCTGGTGGTGG	975
Ddb	361	GAATCGGAGGACTCGGGTGGGGAGGACGTCCGGGTGGGATGAAGTCTGGTGGTGG	420
Oy	976	"TCGTAACT"TAGGAGTGCATGCATCCTCCAGCATCTCAACTCCGCTGTGTCAC"TGTCGTC	1035
Ddb	421	TCGTAACTTAGGAGTGCATGCATCCTCCAGCATCTCAACTCCGCTGTGTCAC"TGTCGTC	480
Oy	1036	AGACTTCGCGACCAATTAGGAATGAGATCCGTGAGATCCTTCCATCTTCTTGAAGTCGC	1095
Ddb	481	AGACTTCGCGACCAATTAGGAATGAGATCCGTGAGATCCTTCCATCTTCTTGAAGTCGC	540
Oy	1096	CTTTAGGCTGGCTCGCAGGT-AGAGGGTGGGGGTGGTGGCTGTCA	1142
Ddb	541	CTTTAGGCTGGCTCGCAGGTAAAAGGTTGGGGGTGGTGGCTGTAA	588

RESULT	12
AV690060	
LOCUS	AV690060 617 bp mRNA EST
DEFINITION	AV690060 GK Homo sapiens cDNA clone GKCAA04 5', mRNA sequence

RECORD KEYWORDS	AB500600.1	GI:10292323
SOURCE	EST.	
ORGANISM	human	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoide; Homo.	
REFERENCE	1 (bases 1 to 617)	
AUTHORS	Wu, T., Qian, H., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H., Xu, X., Li, N., Peng, L., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z. and Han, Z.	
TITLE	Homo sapiens cDNA GK- clones	
JOURNAL	Unpublished (2000)	

Dr. Xinyang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzgchgc.sh.cn
This clone is available at CHGC in Shanghai.

```

FEATURES
  source
    1. .617
      Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="GKBA04"
        /clone_lib="GKC"
        /tissue_type="hepatocellular carcinoma"
        /dev_stage="Adult"
        /lab_host="SOLR"
        /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
        XhoI"

```

Query Match 45.48; Score 557; DB 38; Length 617;

[illegible]


```
source
1. 548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2700722"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT730-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco R1;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 157 a 171 c 98 g 122 t
ORIGIN

Query Match 44.4%; Score 544.8; DB 88; Length 548;
Best Local Similarity 99.6%; Pred. No. 2.8e-128;
Matches 546; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 660 CGAGGCACAGCCAGCTGGGCTAGCGCCGCTGGGAAGGAGAGCGTCTTAATTTATTTTC 719
Db 548 CGAGGCACAGCCAGCTGGGCTAGCGCCGCTGGGAAGGAGAGCGTCTTAATTTATTTTC 489
QY 720 TTATTCCTCCTAATTAATTTATATGATTTATATGATCTCTCTCTAGGTGATGAGATG 779
Db 488 TTATTCCTCCTAATTAATTTATATGATTTATATGATCTCTCTCTAGGTGATGAGATG 429
QY 780 TTATTCCTCCTAATTTATTAATTTATGCAAGGCTGAGATGTTCCCTCTCTGCTGAATG 839
Db 428 TTATTCCTCCTAATTTATTAATTTATGCAAGGCTGAGATGTTCCCTCTCTGCTGAATG 369
QY 840 CAGGCTCTCTGATTTATTTAGCTTTTGTGGACTGGTGGAGCAGGACACCTGGAATCT 899
Db 368 CAGGCTCTCTGATTTATTTAGCTTTTGTGGACTGGTGGAGCAGGACACCTGGAATCT 309
QY 900 CGCCAAAGTAGGAGAAATGGGAGGACTCGGCTGGGGAGGAGACGTCCTGCTCGGAT 959
Db 308 CGCCAAAGTAGGAGAAATGGGAGGACTCGGCTGGGGAGGAGACGTCCTGCTCGGAT 249
QY 960 GAAGTCTGCTGCTGGGTGCTAGTTTAGGAGTGACTGCTCTCCAGCATCTCAACTCC 1019
Db 248 GAAGTCTGCTGCTGGGTGCTAGTTTAGGAGTGACTGCTCTCCAGCATCTCAACTCC 189
QY 1020 GTCTGTCTACTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCCTGAGATCTCTTC 1079
Db 188 GTCTGTCTACTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCCTGAGATCTCTTC 129
QY 1080 ATCTTTCTGAGCTCCCTTTAGGGTGGCTGGCAGGTAGAGGTTGGGGTTGGTGGGCTG 1139
Db 128 ATCTTTCTGAGCTCCCTTTAGGGTGGCTGGCAGGTAGAGGTTGGGGTTGGTGGGCTG 69
QY 1140 TCACGAGGCACTGTGAGATCCCTAGTATGTTCTGTGAACACAAATAAATGATTATA 1199
Db 68 TCACGAGGCACTGTGAGATCCCTAGTATGTTCTGTGAACACAAATAAATGATTATA 9
QY 1200 CTGTCAAA 1207
Db 8 CTGTCAAA 1

RESULT 14
BE878459
LOCUS
DEFINITION 601487866F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890248 5',
mRNA sequence
ACCESSION BE878459
```

```
VERSION BE878459.1 GI:10327235
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1070)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/BTP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MCC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM660 row: c column: 17
High quality sequence stop: 645.
FEATURES
source
1..1070
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3890248"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: Not 1;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 276 a 324 c 302 g 168 t
ORIGIN

Query Match 44.0%; Score 540.8; DB 137; Length 1070;
Best Local Similarity 95.6%; Pred. No. 3.3e-127;
Matches 675; Conservative 0; Mismatches 17; Indels 14; Gaps 11;

QY 151 CGCCCCAGCGCTCTTCGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCC 208
Db 1 CGCCCCAGCGCTCTTCGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCC 60
QY 209 TGGTCCGGCGCCAG-CTGCCAGTCGAGGACCGAAGCCAGCCAA-AGGCTTCTCTTCT 266
Db 61 TGGTCCGGCGCCAGTCCTGCCAGTCGAGGACCGAAGCCAGCCAAAGAGGCTTCTCTTCT 120
QY 267 GCTGC-TCACCATCCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCC 325
Db 121 GCTGCATCACCATCCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCC 180
QY 326 TGCTCCAGAGGAC-GCCCTTAACGCCGATCCTCTGGC-GCCACCCCTGTGTCCCGCG 383
Db 181 TGCTCCAGAGGAGTGCCTTAACGCCGATCCTCTGGCTGGCCGCCCTGTGTCCCGCG 240
QY 384 CCTCAGGCGCTTAACTGTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGCAGCTT 443
Db 241 CCTCAGGCGCTTAACTGTGACTTCGGAGCCCTCGGACTACGCTTGGACCTTAGCAGCTT 300
QY 444 CCTCCAGCAACACCGCGCGCTTCTAACTGTGACTTCGCCGCTCCGCCAGCTCCCAAGAAATC 503
Db 301 CCTCCAGCAACACCGCGCGCTTCTAACTGTGACTTCGCCGCTCCGCCAGCAATCC 360
QY 504 GAAACACCAAGAAACACACAGGCTACCTGCTGGCCAGAGCGTATCCCAACTGGGA 563
Db 361 GAAACACCAAGAAACACACAGGCTACCTGCTGGCCAGAGCGTATCCCAACTGGGA 420
QY 564 CTTCCGAGGCAACTTGAAGTACAGACACTACAGCGGAGACGACCCCGGTGCTTGAAGCG 623
Db 421 CTTCCGAGGCAACTTGAAGTACAGACACTACAGCGGAGACGACCCCGGTGCTTGAAGCG 480
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 22:04:48 : Search time 2136.76 Seconds
(without alignments)
509.900 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468
Perfect score: 258
Sequence: 1 GTCCGGCCAGTCAGT.....AGCAACACCGCGCGCTTC 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 12325308 seqs, 2111498085 residues
Total number of hits satisfying chosen parameters: 24650616

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
 - 2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
 - 3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
 - 4: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
 - 6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
 - 7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
 - 8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
 - 9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
 - 10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
 - 11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
 - 12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
 - 13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
 - 14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
 - 15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
 - 16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
 - 17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
 - 18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
 - 19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
 - 20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
 - 21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
 - 22: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
 - 23: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
 - 24: /cgn2_6/ptodata/2/pna/US097_COMB.seq.*
 - 25: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
 - 26: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
 - 27: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
 - 28: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
 - 29: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
 - 30: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
 - 31: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
 - 32: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
 - 33: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
 - 34: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
 - 35: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
 - 36: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
 - 37: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
 - 38: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
 - 39: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
 - 40: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
 - 41: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
 - 42: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
 - 43: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*

- 44: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
- 45: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
- 46: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
- 47: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
- 48: /cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
- 49: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
- 50: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
- 51: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
- 52: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
- 53: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
- 54: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
- 55: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
- 56: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
1	258	100.0	453	19	US-09-528-409-88514	Sequence 88514, A
2	258	100.0	506	24	US-09-726-805-429	Sequence 429, App
3	258	100.0	521	22	US-09-644-869-356	Sequence 356, App
c	258	100.0	521	44	US-60-196-718-2355	Sequence 2355, App
	258	100.0	529	23	US-09-699-998-4785	Sequence 4785, App
	258	100.0	617	17	US-09-396-885-430	Sequence 430, App
	258	100.0	657	19	US-09-522-303-1025	Sequence 1025, App
7	258	100.0	673	23	US-09-698-010-14935	Sequence 14935, A
8	258	100.0	673	23	US-09-699-998-9797	Sequence 9797, App
9	258	100.0	673	24	US-09-710-286-3772	Sequence 3772, App
10	258	100.0	673	24	US-09-716-473-2500	Sequence 2500, App
11	258	100.0	673	24	US-09-721-588-4548	Sequence 4548, App
12	258	100.0	673	24	US-09-726-787-2982	Sequence 2982, App
13	258	100.0	673	24	US-09-726-787-2982	Sequence 1778, App
14	258	100.0	706	23	US-09-699-998-10469	Sequence 10469, A
15	258	100.0	803	23	US-09-652-109-10087	Sequence 9303, App
16	258	100.0	888	23	US-09-652-126-9303	Sequence 9065, App
17	258	100.0	888	23	US-09-652-126-9303	Sequence 2323, App
18	258	100.0	888	23	US-09-652-126-9303	Sequence 1375, App
19	258	100.0	888	24	US-09-716-473-2323	Sequence 4181, App
20	258	100.0	888	24	US-09-716-473-2323	Sequence 2042, App
21	258	100.0	888	24	US-09-721-588-4181	Sequence 1977, App
22	258	100.0	888	24	US-09-726-172-2642	Sequence 2896, App
23	258	100.0	888	24	US-09-726-805-1977	Sequence 9, Appli
24	258	100.0	888	55	US-09-726-805-1977	Sequence 9, Appli
25	258	100.0	1228	11	US-08-799-910-9	Sequence 9, Appli
26	258	100.0	1228	12	US-08-825-486-9	Sequence 9, Appli
27	258	100.0	1228	12	US-08-826-248-9	Sequence 5, Appli
28	258	100.0	1228	12	US-08-870-434-5	Sequence 9, Appli
29	258	100.0	1228	13	US-08-925-588-9	Sequence 9, Appli
30	258	100.0	1228	17	US-09-372-044-9	Sequence 46, Appli
31	258	100.0	1308	1	PCT-US00-15136-46	Sequence 2393, App
32	258	100.0	1316	17	US-09-338-425-2393	Sequence 9722, App
33	258	100.0	1316	22	US-09-649-162-9722	Sequence 9454, App
34	258	100.0	1316	23	US-09-652-109-9454	Sequence 10214, A
35	258	100.0	1316	23	US-09-652-128-10214	Sequence 9119, App
36	258	100.0	1316	23	US-09-652-128-10214	Sequence 9799, App
37	258	100.0	1316	23	US-09-699-998-9799	Sequence 2658, App
38	258	100.0	1316	24	US-09-710-280-2658	Sequence 2888, App
39	258	100.0	1316	24	US-09-716-473-2888	Sequence 3900, App
40	258	100.0	1316	24	US-09-721-588-3900	Sequence 2410, App
41	258	100.0	1316	24	US-09-726-171-2410	Sequence 2393, App
42	258	100.0	1316	24	US-09-726-790-2393	Sequence 1703, App
43	258	100.0	1316	24	US-09-726-791-1703	Sequence 1978, App
44	258	100.0	1316	24	US-09-726-805-1978	Sequence 2984, App
45	258	100.0	1316	55	US-09-726-787-2984	Sequence 2984, App

ALIGNMENTS

RESULT 1
US-09-528-409-88514

Query Match	100.0%	Score 258;	DB 19;	Length 453;
Best Local Similarity	100.0%;	Pred. No. 7.4e-52;		
Matches 258; Conservative	0;	Mismatches	0;	Indels
Gaps	0;			
aa	1	GTCCGGCCCCAGCTGCCAGTCGACGAGAACCCAGCCAGCAAAAGAGCGTCTCTCTTTCTCATTCTGCTTG	60	
bb	51	gtcggcgccagctgcgactcgagggaacgcgaccaccaagcgcacaagaagtctctcttcttgctg	110	
yy	61	CTACCATCTCTTCTTGCACGATCTGTATGGCTGAAGAGGGTGTCCCGCGCCCTCGCT	120	
bb	111	ctaccatcgtctcttcgcagatccctgatggctgaagagggtgtgccggcgccctgcct	170	
yy	121	CCAGAGAGCGCCCTTAACGCCCATCTCCCTGGCGCCCACCCTGTCTGCCCGTCTCTCGAG	180	
bb	171	cetagaggcgcccctaagcgcgcatccttgcgcgccccaccttggtcccccgctccgag	230	
yy	181	CCCTTTTAATCTGACTTCGAGAGCCCTCGGACTACGCTCTGGACCTCAGACATTGCTCCAG	240	
bb	231	cccccttaaactgacttcggagccctcggagctacgctctggacctcageactttcctccag	290	
y	241	CAACACCGCGCGCCTTC	258	
	291	caaacacccggcgccctc	308	

RESULT 2
US-09-726-805-429

```

: Sequence 429, Application US/09726805
: GENERAL INFORMATION:
: APPLICANT: Gearing, David P.
: APPLICANT: Holtzman, Douglas A.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600.2017-001
: CURRENT APPLICATION NUMBER: US/09/726,805
: PRIOR FILING DATE: 2000-11-30
: PRIOR APPLICATION NUMBER: 60/168,140
: PRIOR FILING DATE: 1999-11-30
: NUMBER OF SEQ ID NOS: 2158
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 429
: LENGTH: 506
: TYPE: DNA
: ORGANISM: Homo sapiens

```

Query Match	100.0%;	Score 258;	IDB 24;	Length 506;
Best local Similarity	100.0%;	pred. No. 7.5e-52;		
Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTCCGGCGCCAGTGCAGTCGAGGAAACCGAACCCGACCAAGGCTTCTCTTCTGCTG 60		
DB	138	gtccgcgccagctgcagtcgaggaaacgaaacccgcaaaaggctctcttctgctg 197		
QY	61	CTCACCATCGTCTTCTGCCAGATCTCTGATGGGTGAAGAGGGTGTGCGCGCGCCCTTGCCT 120		
DB	198	ctcaccatcgctctctgacagatcctgtaggtgtaagagggtgtgctgagagggtgtgctgagggccctgect 257		
QY	121	CCAGGAGCGCCCTTAACCGCGCATCCCTTGGCGCCACACCCCTGTGTCGCCCGTCCCTCGAG 180		
DB	258	ccagaggacgccccaaacgcgcatcccttgcgcccaacccctgltgccccccgtccctcgag 317		
QY	181	CCCTTTAATCTGACTTCGAGGCGCTCGGACTACGCTTCTGGAGCTCAGCACTTCTCTCCAG 240		
DB	318	ccctttcaactgacttcgagccctcggaactacgctcggaactcgacacttctccctccag 377		
QY	241	CAACACCCGGCGCGCTTC 258		
DB	378	caacacccggcgcccttc 395		

```

US-09-644-869-356
; Sequence 356, Application US/09644869
; GENERAL INFORMATION:
; APPLICANT: ShYJan, Andrew W.
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Monahan, John
; APPLICANT: Richardson, Jennifer
; TITLE OF INVENTION: NOVEL NUCLEIC AC
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.1182-001
; CURRENT APPLICATION NUMBER: US/09/644869
; CURRENT FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: 60/151,063
; PRIOR FILING DATE: 1999-08-27
; NUMBER OF SEQ ID NOS: 9708
; SOFTWARE: FastSeq for Windows Versio
; SEQ ID NO 356
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-644-869-356

```

	Query Match	100.0%;	Score 258;	DB 22;	Length 521;
	Best Local Similarity	100.0%;	Pred. No. 7.5e-52;		
	Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
0y	1	GTCCGGCGCCAGCTGGCCAGTCGAGGAACCGAACCCAGCCAAAGAGCGTCTCTTTTCTGCTG	60		
0b	211	gtccggcgccagctgccagtcgaaggaaacccagccaaagagcttctcttctctcgtg	270		
0y	61	CTACCAATCGCTTCTGCCAGATCTTGATGGCTGAAGAGGTGTGCCGGCGCTCGCT	120		
0b	271	ctaccatcgctctctgcagatcctcgtatggctgaaggagggtgtgcggcgccctgcct	330		
0y	121	CCAGAGAGCGCCCTAACGCCGCATCCCTGGCGCCACCCCTGTGTCGCCCGTCTCTCGAG	180		
0b	331	ccagaggagcgccctaagcgcgcatcccttggcgccaccctctgtgtcccccgctctctcga	390		
0y	181	CCCTTTAAATCTGACTTTCGGAGCGCTCGGACATACGCTCTGGACCTTACGACTTTCTCCAG	240		
0b	391	ccctttaactcgacltcggagccctcggactacgctctggacctcagcaacttctctccag	450		

Best Local Similarity 100.0%; Pred. No. 7.50-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGGAACCCAGCCCAAGAGGCTTCTCTTCTCTG 60
Db 182 gtccggcgccagctgccagtcgaggaacccagcccaaaaggctctctctctgctg 241
Qy 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGCTGAAGAGGCTGTGCCGGCGCCCTGCGCT 120
Db 242 ctccaccatcgctctctgccagatccctgatggtgaaaggaggtatgcccggcgccctgcct 301
Qy 121 CCAGAGGACGCCCTTAACGCGCATCCCTGGCGCCCAACCCCTGTGTCCCGCTCTCCAG 180
Db 302 ccagaggagcgccctaaacgcgcgcctcctggtgcccacccctgtgtcccccgtcctcag 361
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACACTTCTCCAG 240
Db 362 ccctttaatctgacttcggagccctcggactacgtctgtggacctcagcaacttctctccag 421
Qy 241 CAACACCGCGCGCCTTC 258
Db 422 caacaccggcgcccttc 439

RESULT 6

US-09-396-885-430
; Sequence 430, Application US/09396885
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
; TITLE OF INVENTION: HUMAN ESOPHAGUS LIBRARY
; FILE REFERENCE: MLN98-48PM
; CURRENT APPLICATION NUMBER: US/09/396,885
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: 60/100,459
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/106,444
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 60/126,842
; PRIOR FILING DATE: 1999-03-30
; NUMBER OF SEQ ID NOS: 5360
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 430
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(617)
; OTHER INFORMATION: n = A,T,C or G
US-09-396-885-430

Query Match 100.0%; Score 258; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 7.6e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGGAACCCAGCCCAAGAGGCTTCTCTTCTGCTG 60
Db 106 gtccggcgccagctgccagtcgaggaacccagcccaaaaggctctctctctgctg 165
Qy 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGCTGAAGAGGCTGTGCCGGCGCCCTGCGCT 120
Db 166 ctccaccatcgctctctgccagatccctgatggtgaaaggaggtatgcccggcgccctgcct 225
Qy 121 CCAGAGGACGCCCTTAACGCGCATCCCTGGCGCCCAACCCCTGTGTCCCGCTCTCCAG 180
Db 226 ccagaggagcgccctaaacgcgcgcctcctggtgcccacccctgtgtcccccgtcctcag 285
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACACTTCTCCAG 240
Db 286 ccctttaatctgacttcggagccctcggactacgtctgtggacctcagcaacttctctccag 345

Query Match 100.0%; Score 258; DB 23; Length 529;

Qy 241 CAACACCGCGCGCCTTC 258
Db 451 caacaccggcgcccttc 468
RESULT 4
US-60-196-718-2355/C
; Sequence 2355, Application US/60196718
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000456
; CURRENT APPLICATION NUMBER: US/60/196,718
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2355
; LENGTH: 521
; TYPE: DNA
; ORGANISM: HUMAN
US-60-196-718-2355

Query Match 100.0%; Score 258; DB 44; Length 521;
Best Local Similarity 100.0%; Pred. No. 7.5e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGGAACCCAGCCCAAGAGGCTTCTCTTCTGCTG 60
Db 400 gtccggcgccagctgccagtcgaggaacccagcccaaaaggctctctctctgctg 341
Qy 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGCTGAAGAGGCTGTGCCGGCGCCCTGCGCT 120
Db 340 CTCACCATCGTCTTCTGCCAGATCCCTGATGCTGAAGAGGCTGTGCCGGCGCCCTGCGCT 281
Qy 121 CCAGAGGACGCCCTTAACGCGCATCCCTGGCGCCCAACCCCTGTGTCCCGCTCTCCAG 180
Db 280 CCAGAGGACGCCCTTAACGCGCATCCCTGGCGCCCAACCCCTGTGTCCCGCTCTCCAG 221
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACACTTCTCCAG 240
Db 220 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACACTTCTCCAG 161
Qy 241 CAACACCGCGCGCCTTC 258
Db 160 caacaccggcgcccttc 143

RESULT 5

US-09-699-998-4785
; Sequence 4785, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4785
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-699-998-4785

Search completed: January 30, 2001, 21:28:30
Job time: 19992 sec

RESULT 9

US-09-206-537-7/c
; Sequence 7, Application US/09206537
; Patent No. 6130052
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coullie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/206,537
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,998
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4524 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 174...1433
; US-09-206-537-7

Query Match 3.4%; Score 41.2; DB 3; Length 4524;
Best Local Similarity 53.8%; Pred. No. 0.18;
Matches 85; Conservative 0; Mismatches 73; Indels 0; Caps 0;
QY 35 CCATCTGACGAGCCCGGCGCCCTCCACCATCCGCGGAGCCCGCGCGGCGTCCG 94
Db 697 CCGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 638
QY 95 GTCTGTGATCTTACCTTCGACCTCTCCGCGAGCCCGGCGGCGGCGGCGGCG 154
Db 637 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 578
QY 155 CCAGCGCTCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 192
Db 577 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540

RESULT 10

US-09-128-155-16

; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANCO-77 RELATED PROTEIN FAMILY
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
; US-09-128-155-16

Query Match 3.3%; Score 40.6; DB 3; Length 152331;
Best Local Similarity 44.3%; Pred. No. 1.2;
Matches 166; Conservative 0; Mismatches 209; Indels 0; Gaps 0;
QY 25 CCGACATGACCATCTGCGAGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 84
Db 21931 ccccccccgcccccccgcccccccgcccccccgcccccccgcccccccg 21990
QY 85 CGGGGCTCGGTCTGAGATCTTACCTTCGACCTCTCCGCGAGCCCGCGGCGGCGG 144
Db 21991 cgc 22050
QY 145 GCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 204
Db 22051 ccccccccgcccccccgcccccccgcccccccgcccccccgcccccccg 22110
QY 205 CGAGTGGTCCGCGCGCGCGCTGCGAGGAGCGGCGGCGGCGGCGGCGGCGGCG 264
Db 22111 cgc 22170
QY 265 CTGCTGCTCACCATCTCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 324
Db 22171 cgc 22230
QY 325 CTGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 384
Db 22231 cagccccgacgc 22290
QY 385 CTCGAGCGCCCTTAAT 399
Db 22291 gcagtcgcgtataat 22305

RESULT 11

US-08-087-007-2
; Sequence 2, Application US/08087007
; Patent No. 5705732
; Patent No. 5705732 5684223
; GENERAL INFORMATION:
; APPLICANT: Sims, Peter J.
; APPLICANT: Bothwell, Alfred L.M.
; APPLICANT: Elliott, Eileen A.
; APPLICANT: Flavel, Richard A.
; APPLICANT: Madri, Joseph
; APPLICANT: Rollins, Scott
; APPLICANT: Bell, Leonard
; APPLICANT: Squinto, Stephen

	Query Match	3.4%	Score 41.2;	DB 2;	Length 4524;
	Best Local Similarity	53.8%;	Pred. No. 0.18;		
	Matches 85;	Conservative 0;	Mismatches 73;	Indels 0;	Gaps
QY	35	CCATCTGTGAGGCCCCGACCCCGCGCCCTCTCCACCATCTCCGGGACCCCGGGGGCTCCG	94		
Db	697	CCGCCCCCGGGCCGCGCACCCGCGCTCTCCCGGGGGCGCGCGCGCCCGCGCCCGCGCG			
QY	95	GTCTGTGAGATTTTACCTTTTCAGACCTCTCTCCCGGAGCCCGGAGCGGCGCTCTCGGGGGCC	154		
Db	637	CCGCCCCGGGCGCCACCGCCCCCGGGCGCTCTGTGGGCGCGCGCGCGCGCGCGCGCG	578		
QY	155	CCAGGCGCTTCTCGGGGACCCGAAAGCGACCGCCGACGG	192		
Db	577	CCCGCGCGCGCCGCGCGCGCCCATGAGCGCGAGCGACGG	540		

;; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;; NUMBER OF SEQUENCES: 52
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: 1800 Diagonal Road, Suite 500
;; CITY: Alexandria
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22313-0299
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/232,463
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US/07/935,313
;; FILING DATE:
;; APPLICATION NUMBER: EP 91 114 300.6
;; FILING DATE: 26-AUG-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703)836-9300
;; TELEFAX: (703)683-4109
;; TELEX: 899149
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7218 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; CLONE: pTZ9pt-Fls
;; US-08-232-463-14

Query Match 3.6%; Score 43.8; DB 1; Length 7218;
Best Local Similarity 2.4%; Pred. No. 0.046;
Matches 9; Conservative 214; Mismatches 156; Indels 0; Gaps 0;

OY 848 TTGGTATTATTACAGCTTTGTGGACTGTGGAACGACACACCTCGGAACCTCGGCAAG 907
Db 1441 TTTGGTACRR 1382
OY 908 TAGGACAAATCGGAGGACTCGGTGGGGAGGAGCTCCCGCTGGGATGAAGTCTG 967
Db 1381 RRR 1322
OY 968 GTGGTGGCTGAAGTTAGGAGGTGACTGCTCCAGCATCTCACTCGGTCTGTCT 1027
Db 1321 RRR 1262
OY 1028 ACTGCTGAGACTTCGCGGACCATAGGAATGAGATCCGTGAGATCTTCCATCTCTT 1087
Db 1261 RRR 1202
OY 1088 CAAGTCGCCTTTAGGTGGCTCCGAGGTAGAGGTTGGGGTTGGGTGGCTGCACGGAG 1147
Db 1201 RRR 1142
OY 1148 CGACTGTGAGATCGCTAGTATGTTCTGTGAACAAATAAATTGATTACTGTCAAA 1207
Db 1141 RRR 1082
OY 1208 AAAAAAATAAATCGC 1226
Db 1081 RRRRRRRRRRRRRATCG 1063

RESULT 6
US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EPO and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P.
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8438 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: pseudorabies virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 622..6495
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1099, "g")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1267, "t")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1381, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1566, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(7010, "g")
; US-07-945-283-1

Query Match 3.4%; Score 42; DB 1; Length 8438;
Best Local Similarity 47.8%; Pred. No. 0.15;
Matches 153; Conservative 0; Mismatches 165; Indels 2; Gaps 1;
OY 56 CGGCCCCCTCCACCATCCCGGACCCCGGGGCTCGGTCTGAGATCTTCACCTTC- 114
Db 6455 CGCGGGGTCCCCGGTCTCCATCAGGGCTCCGAGCAGCGTCTGCTCGTCTCG 6396

Db 421 tacgctctggagctcagcaacttctctccagcaacacccggcgccttcttaactgtgactc 480
Qy 481 CCCGACCTCCCAAAAGAAATCCGAAACACCAAGAAACACCAAGCGGTACCTGGTGGC 540
Db 481 cccgcaactccccaaaagatccgaaaacacacaaagaaacacccagcgctaccctggcg 540
Qy 541 CGAGACCGTATCCCACTGGGACTCCAGCCGCTCCAGCCGCTCCAGCCGCTCCAGCCG 600
Db 541 cgagagcglatcccccaactgggaactccgagggcaactlgaaactcaagacactcaagcgga 600
Qy 601 GAGCCACCCCGTCTTGGGCGGACCGAGCGCACAGACCCGAGCGCATAGAGACC 660
Db 601 gagcgaccccggtcttgagcgagccagcgagcgagcgagcgagcgagcgagcgagcgag 660
Qy 661 GAGGCACAGCCAGCTGGGCTAGGCCCGCTGGGAGGAGAGCGTCTTAATTTATTTCT 720
Db 661 gagggcagcccgagctggggcgtagcccggtgggagagcgagcgagcgagcgagcgag 720
Qy 721 TATTGCTCCTAATTAATTTATATATATATATATATATATATATATATATATATAT 780
Db 721 tatgctcccaatlaatatattatattatattatattatattatattatattatattat 780
Qy 781 GTACGTATATATTTTAACTATGCAAGGCTGTGAGATGTTCCCTGCTGCTAAATGC 840
Db 781 gtacgtatattattttaaacttatgcaaggggtgagatgttccctctctgtgtaaatgc 840
Qy 841 AGGTCTCTTGGTATTATTAGCTTTGGGACTGTGGAGTGGTGAAGCAGCACCTTGAACATGC 900
Db 841 aggtctcttgglatlattatagacttctgggagactgggtgggagagagagagagagag 900
Qy 901 GCAAGTAGAGAGAAATAGGAGGAGCTCGGCTGGGAGGAGGAGCTCGGCTCGGATG 960
Db 901 ggcagagtagagagagaaatgggagagactcgggtgggagagagagagagagagagagag 960
Qy 961 AGCTCTGGTGGTGGTCTAAGCTTAGGAGGTGACTGCTCCTCAGCATCTCAACTCGG 1020
Db 961 agctctgggtgggtgggtgag 1020
Qy 1021 TGTGCTACTGTGAGACTTCGGGGACCATTAGGAATGAGATCCGTGAGATCCTTCCA 1080
Db 1021 tctgctactgtgagagacttcggcgagacattagagaaagagatccglagagatccttcca 1080
Qy 1081 TCTTCTGAAGTGGCTTTAGGCTGGCTCGAGGTAGAGGTGGGGTGGTGGGCTGT 1140
Db 1081 tctcttgaagtcgcctttaggtgaggtgagaggtgagaggtgagaggtgagaggtgag 1140
Qy 1141 CACGAGGAGCTGTGAGATCCGCTAGTATGTTCTGTAACACAAATAAAATTTGATTTAC 1200
Db 1141 cacygagcagctgtcgagatccctagtagtattctgtgaacacacacacacacacacac 1200
Qy 1201 TGTCAAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 1228
Db 1201 tgtcaaaaaaaaaaaaaaaaaaactcgag 1228

RESULT 4

US-08-232-463-14

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

;

;

;

;

;

;

;

;

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232,463

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER: US/07/935,313

; APPLICATION NUMBER: US/07/935,313

; FILING DATE: EP 91 114 300.6

; FILING DATE: 26-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7218 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: pTZgpl-F15

; US-08-232-463-14

Query Match 4.1%; Score 49.8; DB 1; Length 7218;

Best Local Similarity 1.6%; Pred. No. 0.0012;

Matches 6; Conservative 220; Mismatches 147; Indels 0; Gaps 0;

Qy 97 CCTGAGATCTTACCTTCGACCTCTCCGAGCGCCGAGCGCCCTGCGGGCGGCC 156
Db 1059 CTTGCGATTT 1118
Qy 157 AGGCCTCTCGGGCACCAAGCGAGCGAGGTTCTCTACCTCGAGTGTCTCGG 216
Db 1119 YY 1178
Qy 217 CGCCAGCTGCCAGTCGAGAACCCAGCCAGCCAAAGGCTTCTTCTGCTGCTACC 276
Db 1179 YY 1238
Qy 277 ATGCTCTCTCCAGATCCTGATGCTGAAGAGGTGTGCGGGCGCCCTGCCACAG 336
Db 1239 YY 1298
Qy 337 GAGCGCCCTAACGCGCATCCGCGCCACCCCTGTGTCGCCCGTCTCGAGCCCTTT 396
Db 1299 YY 1358
Qy 397 AATCTGACTCGAGCCCTCGAGTACGCTCTGGACCTCAGACTTTCTCTCAGCAAC 456
Db 1359 YY 1418
Qy 457 CCGCGCGCTTCT 469
Db 1419 YYYYYYYYYYYYYY 1431

RESULT 5

US-08-232-463-14/c

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

;

;

;

;

;

;

US-08-826-246-9

Query Match 100.0%; Score 1228; DB 3; Length 1228;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGTCACACTCCAGCTCCAGCCGACCATGACCATCTGTCAGGCGCCCGAGCCCGGCC 60
DB 1 ATGTCACACTCCAGCTCCAGCCGACCATGACCATCTGTCAGGCGCCCGAGCCCGGCC 60

OY 61 CCCTCCACCATCCCGGACCCCGCGGGCTCGGCTCTGAGATCTTTCACCTTCGACCC 120
DB 61 CCCTCCACCATCCCGGACCCCGCGGGCTCGGCTCTGAGATCTTTCACCTTCGACCC 120

OY 121 CTCCTGGAGCTCCAGCGGCGCTGCTGGCGGGCCCGCCAGCCCTCTGCGGGGACCGAAAG 180
DB 121 CTCCTGGAGCTCCAGCGGCGCTGCTGGCGGGCCCGCCAGCCCTCTGCGGGGACCGAAAG 180

OY 181 CGAGCGCGAGGTTCTCTACCTTCGAGTGGTCCGGCGCCAGCTGCGAGTCCGAGAACCG 240
DB 181 CGAGCGCGAGGTTCTCTACCTTCGAGTGGTCCGGCGCCAGCTGCGAGTCCGAGAACCG 240

OY 241 AACCCAGCCAAAAGGCTTCTTCTGCTGCTCACCATCTCTCTGCCAGATCTCTGATG 300
DB 241 AACCCAGCCAAAAGGCTTCTTCTGCTGCTCACCATCTCTCTGCCAGATCTCTGATG 300

OY 301 GCTGAAGAGGTTGCGCGGCGCCCTCCCTCCAGAGGACCGCCCTAACGCGCATCCCTG 360
DB 301 GCTGAAGAGGTTGCGCGGCGCCCTCCCTCCAGAGGACCGCCCTAACGCGCATCCCTG 360

OY 361 CGCGCCACCCCTCTGTCCTCCCGCTCTGAGCCCTTTAATCTGACTTCGGAGCCCTCGAC 420
DB 361 CGCGCCACCCCTCTGTCCTCCCGCTCTGAGCCCTTTAATCTGACTTCGGAGCCCTCGAC 420

OY 421 TACGCTCTGGACCTTCAGCACTTCTCTCCAGCAACACCGCGCGCTTCTTAACCTGACTC 480
DB 421 TACGCTCTGGACCTTCAGCACTTCTCTCCAGCAACACCGCGCGCTTCTTAACCTGACTC 480

OY 481 CCGCACTCCCAAAAGAAATCGAAACCAACCAAGAAACACAGCGCTTCTGTTGGG 540
DB 481 CCGCACTCCCAAAAGAAATCGAAACCAACCAAGAAACACAGCGCTTCTGTTGGG 540

OY 541 CGAGGCTATCCCACTCGGACTTCCGAGGCACTTGAACCTCAGAACACTACAGCGGA 600
DB 541 CGAGGCTATCCCACTCGGACTTCCGAGGCACTTGAACCTCAGAACACTACAGCGGA 600

OY 601 GACCCACCCGCTGCTTGAAGCGGACCGAGCGCACAGACCGAGCGCATAGAGACC 660
DB 601 GACCCACCCGCTGCTTGAAGCGGACCGAGCGCACAGACCGAGCGCATAGAGACC 660

OY 661 GAGGCACAGCCAGCTGGGCTAGGCGCGTGGAGAGGAGCGTCTTAATTTATTTCT 720
DB 661 GAGGCACAGCCAGCTGGGCTAGGCGCGTGGAGAGGAGCGTCTTAATTTATTTCT 720

OY 721 TATTGCTCTTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTCT 780
DB 721 TATTGCTCTTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTCT 780

OY 781 GTACCTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTCT 840
DB 781 GTACCTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTCT 840

OY 841 AGGTCTCTGTTATTTAGCTTTGTTGGACTGTTGGAGTGGTGGAGGACGACCTGGAACCTG 900
DB 841 AGGTCTCTGTTATTTAGCTTTGTTGGACTGTTGGAGTGGTGGAGGACGACCTGGAACCTG 900

OY 901 GCGAAAGTAGGAGAAATGCGGAGGACTTCGGTGGGCGGAGGAGCGTCCCGCTGGGATG 960
DB 901 GCGAAAGTAGGAGAAATGCGGAGGACTTCGGTGGGCGGAGGAGCGTCCCGCTGGGATG 960

OY 961 AAGTCTGCTGGTGGTGAAGTTTAGGAGGTTAGCTTCATCTCTCCAGCATCTCAACTCC 1020
DB 961 AAGTCTGCTGGTGGTGAAGTTTAGGAGGTTAGCTTCATCTCTCCAGCATCTCAACTCC 1020

DB 961 AAGTCTGGTGGTGGTGAAGTTTAGGAGGTTAGCTTCATCTCCAGCATCTCAACTCC 1020
OY 1021 TCTGCTACTGTGTGAGACTTCGGCGGACCATTAGAATGAGATCCGTGAGATCTTCCA 1080
DB 1021 TCTGCTACTGTGTGAGACTTCGGCGGACCATTAGAATGAGATCCGTGAGATCTTCCA 1080
OY 1081 TCTTCTTGAAGTCGCCCTTTAGGCTGCTCGAGGTAGAGGTTGGGGGTTGGTGGCTGT 1140
DB 1081 TCTTCTTGAAGTCGCCCTTTAGGCTGCTCGAGGTAGAGGTTGGGGGTTGGTGGCTGT 1140
OY 1141 CACGGAGGACTGTGTGAGATCCCTTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC 1200
DB 1141 CACGGAGGACTGTGTGAGATCCCTTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC 1200
OY 1201 TGTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1228
DB 1201 TGTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1228

RESULT 2
US-08-944-495-9
; Sequence 9, Application US/08944495
; Patent No. 6087477
; GENERAL INFORMATION:
; APPLICANT: Faib, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA: US/08/944.495
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1228 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...468
; OTHER INFORMATION:
; US-08-944-495-9

Query Match 100.0%; Score 1228; DB 3; Length 1228;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 15:55:18 ; Search time 49.1 seconds
(without alignments)
4030.649 Million cell updates/sec

Title: US-08-799-910-9
Perfect score: 1228
Sequence: 1 ATGTGTCACCTCGCAGCTG.....RAAAAAAAAAAATCGAG 1228

Scoring table: IDENTITY_NUC *
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1228	100.0	1228	3	US-08-826-246-9
2	1228	100.0	1228	3	US-08-944-495-9
3	1228	100.0	1228	3	US-09-126-640-5
4	49.8	4.1	7218	1	US-08-232-463-14
5	43.8	3.6	7218	1	US-08-232-463-14
6	42	3.4	8438	1	US-07-945-283-1
7	41.2	3.4	530	3	US-08-758-662-4
8	41.2	3.4	4524	2	US-08-845-998-7
9	41.2	3.4	4524	3	US-09-206-537-7
10	40.6	3.3	152331	3	US-09-128-155-16
11	40	3.3	2847	3	US-08-087-007-2
12	40	3.3	2847	3	US-08-483-433-2
13	40	3.3	2847	4	PCR-US92-05920-2
14	39.8	3.2	1120	3	US-09-030-613-1
15	38.6	3.1	6354	3	US-09-058-389A-5
16	38.4	3.1	6803	3	US-08-665-259-19
17	38.4	3.1	6803	3	US-08-762-500-19
18	37.6	3.1	170	3	US-09-058-389A-13
19	37.6	3.1	4425	2	US-08-749-169A-1
20	37.6	3.1	4425	2	US-09-130-032A-1
21	37.4	3.0	1420	2	US-08-909-965C-3
22	37.2	3.0	371	1	US-08-664-596B-25
23	37.2	3.0	371	2	US-08-739-775-3
24	37.2	3.0	20235	1	US-07-642-734C-3
25	37.2	3.0	20235	3	US-08-439-009A-3
26	36.6	3.0	1931	2	US-09-130-114-2
27	36.4	3.0	702	1	US-08-458-568A-3
28	36.4	3.0	1841	2	US-08-820-521-1

Sequence 1, Appli
Sequence 5, Appli
Sequence 11, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 213, App
Sequence 11, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-826-246-9
; Sequence 9, Application US/08826246
; Patent No. 6048709
; GENERAL INFORMATION:
; APPLICANT: Faib, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,246
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-078-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1228 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..468
; OTHER INFORMATION:

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 16:08:34 ; Search time 77.57 Seconds

(without alignments)
5947.064 Million cell updates/sec

Title: US-08-799-910-9

Perfect score: 1228

Sequence: 1 ATCTCTCACTCTCGACGCTG.....AAAAAATACTCGAG 1228

Scoring table: IDENTITY_NUC %

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_36:**
1: /cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	DB	ID	Description
1	1228	100.0	1228	21	250711	Nucleotide sequenc
2	1226.4	99.9	1228	18	T94471	Human Fchd605 gene
3	420.8	34.3	427	21	280483	Human colon cancer
4	371.6	30.3	411	20	V86227	EST clone OA17. H
5	167.2	13.6	193	20	V89271	EST clone BY66. H
6	151	12.3	297	15	Q77534	Human genome fragm
7	70.4	5.7	267	15	Q77554	Human genome fragm
8	57.4	4.7	1000	21	A02484	Human colon cancer
9	55.2	4.5	1218	21	A02488	Human colon cancer
10	52.2	4.3	1337	20	Z17263	Human gene express
11	49.6	4.0	1126	21	A02538	Human colon cancer
12	48.2	3.9	53	16	T20789	Human gene signatu

13	47.8	3.9	4228	21	251683	Human cyclic nucle
14	47.2	3.8	114955	20	X53491	Human adenosine A1
15	44.8	3.6	1593	21	A02504	Human colon cancer
16	44.2	3.6	2375	21	A34740	Human adenosine re
17	44.2	3.6	4475	21	A34741	Human adenosine re
18	42.8	3.5	1127	21	A02477	Human colon cancer
19	42.4	3.5	4356	14	Q37543	Human colon cancer
20	42.4	3.5	4356	16	Q95540	Cardiac adenylyl c
21	42	3.4	8438	15	Q73500	Cardiac adenylyl c
22	42	3.4	34094	20	Z30163	DNA encoding Pseud
23	41.8	3.4	1459	21	A02528	Complete nucleotid
24	41.6	3.4	989	21	A02539	Human colon cancer
25	41.4	3.4	1275	11	Q01620	Human colon cancer
26	41.4	3.4	1344	12	Q13318	Chicken beta-actin
27	41.4	3.4	1345	11	Q03062	Expression vector
28	41.2	3.4	1102	20	Z16923	Hybrid promoter of
29	41.2	3.4	4524	20	V33912	Human gene express
30	41	3.3	114955	20	X53491	Nucleotide sequenc
31	40.8	3.3	1966	20	X61220	Human adenosine A1
32	40.6	3.3	1017	20	Z17058	Mouse DNA demethyl
33	40.4	3.3	300	20	Z14922	Human gene express
34	40.4	3.3	2179	15	Q73117	Human gene express
35	40.4	3.3	2232	15	Q77739	Alpha 2, 3-sialyl
36	40	3.3	1523	20	Z17507	Human alpha-2,3-si
37	40	3.3	1523	20	Z17479	Human gene express
38	39.8	3.2	1120	19	V60339	Human gene express
39	39.8	3.2	2188	20	Z77506	cDNA sequence of f
40	39.8	3.2	3198	20	X02974	Human ovarian tumo
41	39.6	3.2	954	20	X03042	Human IL-1ra BAC c
42	39.6	3.2	2681	20	X58429	DNA encoding a hum
43	39.2	3.2	1804	20	X61218	Thermophilus therm
44	38.8	3.2	749	20	Z16153	Human DNA demethyl
45	38.8	3.2	1312	20	Z17264	Human gene express

ALIGNMENTS

RESULT 1	
ID 250711	standard; DNA; 1228 BP.
AC 250711;	
DT 31-MAY-2000	(first entry)
XX	
DE	Nucleotide sequence of human fchd605 gene.
XX	
KW	fchd605 gene; human; cardiovascular disease; oncogenic disorder;
KW	diabetic retinopathy; fibroproliferative disorder; atherosclerosis;
KW	TGF-beta signalling pathway; TGF; Transforming growth factor;
KW	pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;
KW	vascularisation; cytostatic; antidiabetic; ophthalmological; ds.
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..471
FT	/*tag= a
FT	/product= "fchd605 protein"
XX	
PN	WO200006206-A1.
XX	
PD	10-FEB-2000.
XX	
PF	30-JUL-1999; 99WO-US17394.
XX	
PR	30-JUL-1998; 98US-0126640.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Falb DA;
XX	

WPI: 2000-205414/18.
P-PSDB: Y45017.
Identifying substances for ameliorating symptoms of fibroproliferative diseases or oncogenic related disorders.
Examples: Fig 5; 214pp; English.
The patent discloses methods for the treatment and diagnosis of cardiovascular diseases by novel human genes which are differentially expressed in different cardiovascular disease states. Compositions which can modify TCP-beta signalling pathway are identified by screening. These are used therapeutically to treat fibroproliferative and oncogenic disorders, especially TGF (transforming growth factor)-beta related disorders, including diabetic retinopathy, arteriosclerosis, pancreatic cancer, angiogenesis, inflammation, fibrosis, tumour growth and vascularisation. The present sequence is fchd605 gene which is up-regulated in monocytes treated with oxidised LDL (low density lipoprotein) can be used to design cardiovascular disease treatment strategies. Depending on whether the up-regulation has a pathogenic or protective effect treatment methods can be designed to increase or decrease the activity of the protein product of the gene.
Sequence 1228 BP; 265 A; 361 C; 341 G; 261 T; 0 other;

Query Match 100.0%; Score 1228; DB 21; Length 1228;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTGTCACATCTCCAGCTGCCACCGACCATGACCATCTCTGAGAGTTCACCTTCGACCT 120
DB 1 algtgtcaetctcgcagctgcacccgacatgacatcccgagggcccccgagcccccgcc 60
QY 61 CCTTCACCATCCGGAGCCCGGGGGCTCCGCTCGAGATCTTCACCTTCGACCT 120
DB 61 ccttcacacatccgagaccccgggggtccggtccgtgagatcttcacccctcgacct 120
QY 121 CTCGGAGCCCGCAGCGCCCTGCGGGGCCCGCCAGCGCTCTCGGGGACCCGAAAG 180
DB 121 ctccggagcccgagcgccctgcggcgcccgagcgccctgcggcgcccgagcgagaa 180
QY 181 CGAGCGCGAGGTTCTTACCTCGAGTGGTCCGCGCGCAGTCCAGTCCAGGAAACCG 240
DB 181 cgcagcgcgaggtctctacccctcagagtggtccggcgccagctgcagtcgaggaaccg 240
QY 241 AACCCAGCCAAAAGGCTTCTTCTGCTGCTCACCATCGTCTTCTGCGAGATCTCGATG 300
DB 241 aaccgcgcaaaaggct 300
QY 301 GCTGAGAGGGTGTGCGCGCGCCCTGCTCCAGAGAGCGCCCTAACCGCGCATCCCTG 360
DB 301 gctgagaggggtgtgccccgccccctgctccagagagcgccccctaacgccccctcg 360
QY 361 CGCCCGACCCCTGTGCCCCGCTCTCGAGCCCTTTAAATCTGACTTCGAGCCCTCGGAC 420
DB 361 ggcggccacccctgtgtcccccgtccctcgagccctttaaactgacttcggagccctcgac 420
QY 421 TAGCCTCTGACCTGACGACTTTCTCTCCAGCAACCCCGCCCTTCTTAACCTGTCATC 480
DB 421 tagcctctgagacctgagccttctctccgagcaaccccgccctcttaactgtgactc 480
QY 481 CCGCGACTCCCCAAAGAAATCCGAAAACACAAAGAACACAGCGCTAGCTGTGTCGG 540
DB 481 ccgcgaccccccaaaagaatccgaaaaaccccaaaagaacacagggcgctactggcg 540
QY 541 CGACACCGTATCCCCAATCGGACTTCCAGGCACTTGAACCTCAGAACACTACACCGGA 600
DB 541 cgagagcgtatcccccaactgggacttccgaggaacttgaaactcagaaacactacacggga 600
QY 601 GACGCCACCCGGTGTGTTAGGGGGGACCGGAGCGGACAGACCGGCGCATAGAGACC 660
DB 601 gacgccacccgggtgtgttgggggggagccgagcgacagagacgagcgcatagagacc 660

QY 661 GAGGCACAGCCCGAGCTGGGCTAGGCCGCTGGGAAGAGAGAGCGTCTTAATTTATTTCT 720
DB 661 gaggcacagcccgagctgggctagggccggagggagagagcgcglaalclaltcl 720
QY 721 TATTGCTCTTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 780
DB 721 tattgctcttaataataattattattattattattattattattattattattattattatt 780
QY 781 GTACGTAATATTTATTTAACATATATCAAGGCTGTGAGATGTTCCCTCTGCTGTAATG 840
DB 781 gtacgtaataattattattattattattattattattattattattattattattattattatt 840
QY 841 AGGTCCTCTTGGTATTTATTTACGCTTTGTGGGACTGTGGAAGCAGCACACCTTGGAACT 900
DB 841 aggtctcttggtattattattattattattattattattattattattattattattattatt 900
QY 901 GCCAAGTAGGAGAGAAATGGGAGGACTCGGGTGGGGAGGAGCGTCCCGGCTGGGATG 960
DB 901 gccaaagtaggagagaaatgggagggactcgggtggggagagcgtcccggtggatg 960
QY 961 AAGTCTGCTGGTGGTTCGTTAAGTTAGGAGGTGACTGTCATCTCCAGCATCTCAACTCG 1020
DB 961 aagtctggtggtggttcgttaagtttaggggtgactgcatctccagcatctcaactcgg 1020
QY 1021 TCTGTCTACTGTGTGAGACTTCGGCGACCATTAGGAATGAGATCCGTGAGATCTCTTCCA 1080
DB 1021 tctgtctactgtgtgagacttcggcgacctcggcgaccattaggaaatgagatccgtgagatcttcca 1080
QY 1081 TCTTCTTGAAGTCGGCTTTAGGGTGGCTCCGAGGTAGAGGTTGGGGTTCGGGCTGCT 1140
DB 1081 tcttcttgaagtcggcttttaggggtggctcgcgaggttaggggttgggtgggctgct 1140
QY 1141 CACGAGCGCAGCTGTGAGATCGCTAGTATGTTCTGTGAACACAAATAAATTTGATTAT 1200
DB 1141 cacgagcgactgtcgagatcgctagatgcttctgtgaacacaaataaattgattat 1200
QY 1201 TGTCAAAAAAATAAAAAAACTCGAG 1228
DB 1201 tgtcaaaaaaataaaaaaactcgag 1228

RESULT 2
T94471
ID T94471 standard; cDNA; 1228 BP.
AC T94471;
XX
XX
DT 03-MAR-1998 (first entry)
XX
DE Human Fchd605 gene differentially regulated in monocytes.
XX
KW Fchd602 gene; differential expression; monocyte; human;
KW foam cell; cardiovascular disease; atherosclerosis; ischemia;
KW reperfusion; hypertension; restenosis; arterial inflammation;
KW therapy; diagnosis; drug screening; marker; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..471
FT /tag= a
XX
PN WO9730065-A1.
XX
PD 21-AUG-1997.
XX
PF 14-FEB-1997; 97WO-US02291.
XX
PR 13-FEB-1997; 97US-0799910.
PR 16-FEB-1996; 96US-0011787.
XX
PA (MIL-) MILLENNIUM PHARM INC.

XX Falb DA;
PI
XX WPI: 1997-424966/39.
DR P-PSDB; W36006.
XX
XX New genes differentially expressed in cardiovascular disease - used
PT for diagnosis, drug screening and treatment of cardiovascular
PT disease, e.g. atherosclerosis, restenosis, hypertension, etc
XX
XX Claim 1: Fig 5; 163pp; English.
XX
XX Fchd605 is a novel human gene that is up-regulated in monocytes
CC treated with oxidised low density lipoproteins (LDL). Differential
CC display was used to detect genes that are differentially expressed
CC in monocytes treated so as to simulate the conditions under which
CC foam cells develop during atherogenesis. Both fchd605 and fchd602
CC (see T94470) are up-regulated under the disease condition of
CC treatment with oxidised LDL. The fchd605 gene product (see W36006)
CC has sequence similarity to mouse gly96. The discovery of the
CC up-regulation of these 2 genes provides a fingerprint profile.
CC e.g. markers for the study of cardiovascular diseases, including
CC atherosclerosis, ischaemia/reperfusion, hypertension, restenosis,
CC and arterial inflammation. Methods are provided for the diagnosis,
CC monitoring in clinical trials, screening for therapeutically
CC effective compounds, and treatment of cardiovascular diseases based
CC on discoveries regarding the expression patterns of novel genes
CC fchd531 (see T94467), fchd540 (see T94468), fchd545 (see T94469),
CC fchd602 and fchd605.
XX
XX Sequence 1228 BP; 265 A; 362 C; 340 G; 261 T; 0 other;
SQ

Query Match 99.9%; Score 1226.4; DB 18; Length 1228;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1227; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGTGTCACTCTCCAGCTGCCACCCGACCATGACATCTGTGAGGCCCGCCGCGGCC 60
Db 1 atgtgtcactctcgcagctgccacccgacccatgacatctgtgagggcccgcccgcc 60
QY 61 CCTTCCACCATCCCGGACCCCGCGGGGCTCCGTCCTGAGATCTTCACCTTCGACCT 120
Db 61 ccttccacccatcccgagaccccgcg999gctccg99gctccgagatcttcaccttcgaccct 120
QY 121 CTCCCGGAGCCCGACGGCCCTGCGCGGCCGCCCGCCGAGCCCTCTCCGCGCACCCGAAAG 180
Db 121 ctcccgagcccgacggccctgcccggcgccccagcgccctcgcgggacccgaaag 180
QY 181 CGCAGCGCAGGGTCTCTACCTCGAGTGGTCCGCGCGCAGCTGCCAGTCGAGGAACCG 240
Db 181 cgcagccgcagggtctctaccctcagtggtccg99gctccg99gctccgagtcgaggaaccg 240
QY 241 AACCCAGCCAAAAGGCTTCTTCTTCTGCTCACCATCGTCTTCCGCGAGATCCTCGATG 300
Db 241 aacccagccaaaaggctctcttcttctgctcaccatcgctcttcgagatcctcgatg 300
QY 301 CTTGAGAGGCTGTCCGCGCCCTGCTTCTTCTGCTCAGAGGACGCCCTTAACCGCCATCCCTG 360
Db 301 gctgaagagggtgtgcccgcgcccctgctctccagagagcgcgccctaaacgcccacccctg 360
QY 361 GCGGCCACCCCTGTGTCGCCCGCTCGAGCCCTTTAACTGACTTCGAGCCCTCGGAC 420
Db 361 gcgccacccctgtgtcccccgtccctcagggccctttaaactgacttcggagccctcgag 420
QY 421 TAGCCTCTGGACCTCAGCACTTTCTCTCCAGCAACACCTCGCCGCTTCTTAAGTGTGACTC 480
Db 421 tagcctcggactcagcacttctccagcaaacacccgcccgccttctaactgtgacttc 480
QY 481 CCGCGACTCCCAAAAGAAATCGAAAAACCAAGAAACACCAAGAGGCGGTACCTGGTGGC 540
Db 481 ccgcgactccccaaaagaatcggaaaaaccccaagaaacacccaggcgctacctggtgcg 540

QY 541 CGAGAGCGTATCCCAACTCGGACTTCCGAGGCAACTGTGACTCAGAACACTACACGGA 600
Db 541 cgaagagcgatcccccaactgggacttccgaggcaacttgaaactcagaacactacagcga 600
QY 601 GACGCCACCCGCTGCTTGTAGCGCGGACCGAGCGCACAGAGACCGAGCGGCATAGAGACC 660
Db 601 gacgccacccgclgcttgaggcggagcgcacagagacccgagggcgcalagagacc 660
QY 661 GAGCGACAGCCCACTGCGGCTAGGCCCGGTGGGAAGGAGAGCGCTCTTAATTTATTTC 720
Db 661 gaggcacagcccagctgggctagcccggtgggaagagagcgtctcgtaattttcttct 720
QY 721 TATTGCTCTTAATAATTTATATCTATTATTATGTAGCTCCCTCTAGGTGATGAGATGT 780
Db 721 tatgctctccaaataattattatgtattatgtatctctcctcctaggtgatggatgt 780
QY 781 GTACGTAATATTTATTTAACTTATCAAGGGTGTGAGATGTTCCCTCTCTCAATGTC 840
Db 781 gtacgtaattatttttaacttatgcaagggtgtagatgtlccctcclgclglaaatgc 840
QY 841 AGTCTCTTGGTATTTATTGAGCTTTGTGGACTGTGTGAAGCAGGACACCTCGAACTGC 900
Db 841 agtctcttggattttatttattgagctttgtggactggtggaagcagacacccctggaactgc 900
QY 901 GGCARAACTAGGAGAAATGGGAGGACTCGGCTGGGGAGGAGACGTCGCCGCTGGGATG 960
Db 901 ggcaraagtaggagaagaaatgggaggaactcggg99gaggaogtcccgcgtggatg 960
QY 961 AACTCTGCTGGTGGGTCGTAAGTTTAGGAGGTGACTGCTATCCTCAGCATCTCAACTCCG 1020
Db 961 aactctggtgggtgggtcgttaagttaggaggtgactgctcctccagcacltcaacclcg 1020
QY 1021 TCTGCTACTGTGAGACTTCGGCGGACCATTTAGGAATGAGATCCGTGAGATCTTCCTCA 1080
Db 1021 tctgctactgtgagacttcggcgaccattaggaatgagatcgatcgatcgatccttcca 1080
QY 1081 TCTTCTTGAAGTCGCTTTTAGGGTGGCTCGGAGGTAGAGGGTGGGGTGGTGGGCTGT 1140
Db 1081 tcttcttgaagtcgctctttagggctgaggtgaggttaggggtctgggggtctggggctgt 1140
QY 1141 CACGAGCGGACTGTGAGATCCCTCTAGTATGTTCTGTCGAACACAATAAATTTGATTTAC 1200
Db 1141 cacgagcgactgtcagatcgcctaglatgtctcgtgaaacacataaaatltgatttac 1200
QY 1201 TCTCAAAAAAATAAAAAAACTCGAG 1228
Db 1201 tctcaaaaaaataaaaaaactcgag 1228

RESULT 3
280483/c
ID 280483 standard; cDNA; 427 BP.
XX 280483;
AC 280483;
XX
DT 07-APR-2000 (first entry)
XX
XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:567.
XX
XX Human; gene expression product; diagnosis; tumour; colon cancer;
XX colorectal adenocarcinoma; cell line SW480; cell proliferation;
XX cytosolic; sarcoma; breast cancer; neoplasia; dysplasia;
XX hyperplasia; ds.
OS Homo sapiens.
XX
XX WO9964576-A2.
XX
XX 16-DEC-1999.
XX
XX 09-JUN-1999; 99WO-1801062.
XX
XX 10-JUN-1998; 98US-0088801.
PR

XX PA (FARB) BAYER CORP.
XX PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Rushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DW, Lewis ME, Monahan JE;
XX PI Schlegel R;
XX DR WPI; 2000-087220/07.
XX XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX PS Claim 15; Page 356; 469pp; English.
XX XX 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX XX Sequence 427 BP; 115 A; 138 C; 78 G; 96 T; 0 other;

Query Match 34.3%; Score 420.8; DB 21; Length 427;
Best Local Similarity 99.5%; Pred. No. 6.4e-102;
Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 785 GTAATATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTCTGTAATGACGGT 844
DB 424 GTAATATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTCTGTAATGACGGT 365
QY 845 CTCTTGATTTTATGAGCTTTCTGGACTGTTGGAGCAGACACCTGGAATCGGCA 904
DB 364 CTCTTGATTTTATGAGCTTTCTGGACTGTTGGAGCAGACACCTGGAATCGGCA 305
QY 905 AAGTAGGAGAGAAATGGGAGGACTCGGCTGGGGAGGAGACGTCCTGGGTGGGATGAAGT 964
DB 304 AAGTAGGAGAGAAATGGGAGGACTCGGCTGGGGAGGAGACGTCCTGGGTGGGATGAAGT 245
QY 965 CTGGTGGTGGTGGTGAAGTTAGGAGTGACTCATCTCCACCATCTCAACTCCGTCG 1024
DB 244 CTGGTGGTGGTGGTGAAGTTAGGAGTGACTCATCTCCACCATCTCAACTCCGTCG 185
QY 1025 TCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCCATCTT 1084
DB 184 TCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCCATCTT 125
QY 1085 CTTGAAGTCGCCCTTTAGGGTGGCTGGCAGGTAGAGGGTTGGGGTGGTGGCTGTACG 1144
DB 124 CTTGAAGTCGCCCTTTAGGGTGGCTGGCAGGTAGAGGGTTGGGGTGGTGGCTGTACG 65
QY 1145 GACGACTGTGAGATCGCCTAGTATGTTCTGTGACACAAATATAATGATTTACTGTC 1204
DB 64 GACGACTGTGAGATCGCCTAGTATGTTCTGTGACACAAATATAATGATTTACTGTC 5

QY 1205 AAAA 1208
DB 4 AAAA 1

RESULT 4
V86227
ID V86227 standard; cDNA; 411 BP.
XX AC V86227;
XX DT 27-APR-1999 (first entry)

XX EST clone 0417.
XX XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX OS Homo sapiens.
XX XX W09845435-A2.
XX PN 15-OCT-1998.
XX PD 10-APR-1998; 98WO-US06954.
XX PR 10-APR-1997; 97US-0835913.
XX PA (GENY) GENETICS INST INC.
XX XX Agostino MJ, Jacobs K, Lavallie EK, McCoy JM, Merberg D;
PI Racine LA, Spaulding V, Treacy M;
XX WPI; 1999-070076/06.
XX XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX PS Claim 1; Page 165; 633pp; English.
XX CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity,
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX SQ Sequence 411 BP; 100 A; 89 C; 122 G; 99 T; 1 other;
Query Match 30.3%; Score 371.6; DB 20; Length 411;
Best Local Similarity 98.5%; Pred. No. 6.3e-89;
Matches 385; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
QY 525 AGCGGTACCTGGTGGCGGAGAGCGTATCCCAACTGGGACTCCGAGGCAACTTGAATC 584
DB 15 aggcctacgtggtgctgagagcggtatcccccaactggactccgaggaacilgaaclt 74
QY 585 AGAACAACACTACAGCGGAGACGCCACCGGTGCTTTGAGCGGGACCGGACGACAGAGACC 644
DB 75 agaacactacagcgagagcgccaccggtgcttgagggcgagcgagcgagcgagagacc 134
QY 645 GAGGCGCATAGACACCGAGGACAGCCAGCTGGGCTAGGCCCGGTGGG-AGAGAGAGC 703
DB 135 gagcgcatagagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 194
QY 704 GTCGTTAATTTATTTCTTATTTCTCTTAATTAATATTTATATGTTATGTCGTCCTC 763
DB 195 gtcgttaattcttattcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 254
QY 764 CTAGGTGATGGAGATGTGTACGTAAATATTTATTTTAACTTATGCAAGCGGTGTGAGATGTT 823
DB 255 ctagggtgagagatgtgtacgttaatttttatttttttaacttattcgaagggtgtgagatgtt 314

Db	133	CTCAGCAGACACCGCGCGCTTCTTACTGTGTACTCCCGCGACTCCCGCCAAAAGATCCG	74
Qy	505	AAAAACACAAAGAAACACACAGCGCTACCTGGTGGCGAGAGCCTATCCCCA	556
Db	73	AAAAACACAAAGAAACACACAGCGTACCTGGTGGCGAAGCGTATCTCGA	22
RESULT	6		
Q77534			
ID	Q77534	standard; DNA; 297 BP.	
XX	XX	Q77534;	
AC	Q77534;		
XX	XX		
DT	23-SEP-1994	(first entry)	
XX	XX	Human genome fragment.	
DE	XX		
XX	XX	Brain; placenta; bone marrow; genetic analysis; gene mapping;	
KW	XX	detection; homology; human; adrenal tissue; ds.	
KW	XX		
XX	XX	Homo sapiens.	
OS	XX		
PN	WO9401548-A.		
XX	XX		
FD	20-JAN-1994.		
XX	XX		
PF	13-JUL-1993;	93WO-GB01467.	
XX	XX		
PR	13-JUL-1992;	92GB-0014857.	
XX	XX	(MEDI-) MEDICAL RES COUNCIL.	
PA	XX		
PI	Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;		
P1	Sibson DR, Starkey M;		
XX	XX		
DR	WPI: 1994-035056/04.		
XX	XX		
PT	New nucleic acid fragment encoding gene products - can be used		
PT	for genetic analysis and mapping		
XX	XX		
PS	Claim 1: Page 575-576; 616pp: English.		
XX	XX		
CC	Human nucleic acid fragments, isolated from brain, adrenal tissue,		
CC	the placenta or bone marrow comprise any of: (A) a sequence		
CC	selected from (076401-077613), (B) an allelic variation of a		
CC	sequence as described in (A), or (C) a sequence complementary		
CC	to (A) or (B).		
CC	Preferred sequences exhibit no more than 90% homology to a human		
CC	sequence known per se.		
XX	XX		
SQ	Sequence 297 BP; 66 A; 117 C; 61 G; 52 T; 1 other;		
Query Match	12.3%;	Score 151;	DB 15; Length 297;
Best Local Similarity	76.0%;	Pred. No. 1e-30;	
Matches	212; Conservative	0; Mismatches	65; Indels
		2; Gaps	
Qy	1	ATGTGTACTCTCGAGCTGCCACCGACCATGACCATCTGTGAGCGCCCGACCCCGGCC	60
Dh	16	atggaacactctcggtgcactccccatgaccaccclacaggccctaacccccagcc	75
Qy	61	CCCTCCACCATCCGGGACCCGGCGGGGCTCCGGTCTGAGATCTACCTTCGACCCCT	120
Dh	76	ctcaagttactgtctcagcacaata-gctgcgccttgagatctccacctcagacct	134
Qy	121	CTCCGGAGCCGACGAGCGGCCCTTCGCGGGCGCCCGACGCGCTCTCGCGGCGACCGAAG	180
Dh	135	ctccgaagccacagcgccctlgcggcaccaccagctc-ctcgaggaccacaaag	193
Qy	181	CGCAGCCGACGGTTCTTACCCYCGAGTGTCCGGCGCCGAGTGTCGAGGACCG	240
Dh	194	cgaactccaggaatctctacccctlagagtatccgqccagtlgccaaglttaaggaacca	253

[illegible]

RESULT	9	
A02488/C		
ID	A02488	standard; cDNA; 1218 BP.
XX		
XX		
AC	A02488;	
XX		
XX		
DT	19-MAY-2000	(first entry)
XX		
DE	Human colon cancer cell line polynucleotide sequence	SEQ ID NO:2479
XX		
KW	Human; colon cancer; tumour; diagnosis; gene expression product;	
KW	probe; detection; cancerous state; metastasis; identification;	
KW	breast cancer; oestrogen receptor-positive breast cancer; therapy;	
KW	oestrogen receptor-negative breast cancer; lung cancer; ss.	

OS	Homo sapiens.	
XX		
XX	W0958675-A2.	
XX		
XX	18-NOV-1999.	
PD		
XX		
PF	13-MAY-1999;	99WO-US10602.
XX		
XX	14-MAY-1998;	98US-0085426.
PR	15-MAY-1998;	98US-0085537.
PR	15-MAY-1998;	98US-0085696.
PR	21-OCT-1998;	98US-0105234.
PR	21-OCT-1998;	98US-0105877.
XX		
XX	(CHIR) CHIRON CORP.	
PA	(HYSE-) HYSEQ INC.	
PA		
XX		
PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;	
PI	Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;	
PI	Leinson G, Drmanac R, Krkenjakov R, Dickson M, Drmanac S, Labat I;	
PI	Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;	
XX		
XX		
DR	WPI: 2000-126369/11.	
XX		
PT	Polynucleotide library used to determine cancerous states of mammalian	
PT	cells -	
XX		
XX	Claim 1; Page 995-996; 1097pp; English.	
PS		
XX		
CC	A00010 to A02716 represent polynucleotides isolated from CDNA libraries	

constructed from human colon cancer cell lines. The present invention also describes a method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, where detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived. The polynucleotide sequences can be used in a method for detecting differentially expressed genes correlated with a cancerous state of a mammalian cell. The polynucleotides can also be used as probes for detecting and mapping related genes. They can be used in diagnosis and prognosis of diseases and disorders (e.g. identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy). This is particularly for breast cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-negative breast cancer, lung cancer, and colon cancer.

Sequence 1218 BP; 9 A; 31 C; 494 G; 37 T; 647 other;

[illegible]

RESULT	10
Z17263	
ID	Z17263 standard; cDNA; 1337 bp.
XX	
XX	
AC	Z17263;
XX	
DT	12-OCT-1999 (first entry)
XX	
XX	
DE	Human gene expression product cDNA sequence SUB ID NO:4735.
XX	
XX	Human; gene; gene expression product; diagnosis; therapy; probe;
KW	detection; mapping; tissue typing; profiling; forensic; cancer;
KW	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX	
OS	Homo sapiens.
XX	
XX	

CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 1126 BP; 101 A; 102 C; 575 G; 57 T; 291 other;

Query Match 4.0%; Score 49.6; DB 21; Length 1126;
Best Local Similarity 33.2%; Pred. No. 0.0012;
Matches 192; Conservative 0; Mismatches 383; Indels 4; Gaps 1;

QY 13 CGCAGTGGCACCACGACATGACATCTTCGAGGCCCCGACCGCCCGCCCTCCACCATC 72
DB 608 CCGGNNCGCTCCNCAACNCGCGCTCCCGAGNCCCATTCCTCCNCGCCGCGCCNNTCAAG 549
QY 73 CGGGGACCCCGCGGGCTCCGCTCTGAGATCTTCACCTTCGACCTCTCCCGGAGCCC 132
DB 548 CTCGNCCTCCCGCGCGCTTCCTCCGCGGACCTCCGCGCCCGCCCGCGCGCCN 489
QY 133 GCAGCGGCGCTCCCGCGCGCGCGCGCTTCGCGGCGACCGAAAGCGCGCGCAGG 192
DB 488 CCTGCGNCCNCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 428
QY 193 GTTCTACCTGAGTGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252
DB 429 ---CCANTCCCTGCGCTCTTCANCCGCTTCGCTTCGCGNCCNCCNCCCTNCC 373
QY 253 AGGCTTCTCTTCTGCTGCTACCATCGTCTTCGCGGACATCTGATGCTGAAGGGT 312
DB 372 CGCTCCNCTCTCNCNCGGCTACNCCNCCCGCGCGCGCGCGCGCGCGCGCGCGCG 313
QY 313 GTGCGCGCGCGCTGCTCCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
DB 312 CG 253
QY 373 TCCCGCGCGCGCTTAACTGACITCGGAGCGCTTCGGAGCTTCGGACTACGCTGGAC 432
DB 252 NCNNGCG 193
QY 433 CTCGACACTTCTCTCCAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492
DB 192 ANCANATNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 133
QY 493 AAAAAAGATCCGAAAAACACAAACACACACAGCGGTACCTGGTGGCGGAGAGGTATC 552
DB 132 CNTCCG 73
QY 553 CCCAACTGGGACTTCGGAGGCACTTGAACCTCAGAACAC 591
DB 72 NNTNCCNCCNCCCTTNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 34

RESULT .12
T20789
ID T20789 standard; cDNA to mRNA; 53 BP.
XX
AC T20789;
XX
DT 09-JUL-1996 (first entry)
XX
DE Human gene signature HUMGS02013.
XX
KW Gene signature: messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.

XX
PN W09514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUBO/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 723; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(r) as the sole primer. Since the 3'-
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 53 BP; 18 A; 9 C; 9 G; 17 T; 0 other;

Query Match 3.9%; Score 48.2; DB 16; Length 53;
Best Local Similarity 94.3%; Pred. No. 0.00069;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1158 GATCGCCCTAGTATGTTCTGTGAACACAAATAAATTTGATTCTGTCAAAAAA 1210
DB 1 gatcgccctagtatgttctgtgaacacaaataaattgatttactgtctgcaaa 53

RESULT 13
251683
ID 251683 standard; cDNA; 4228 BP.
XX
AC 251683;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human cyclic nucleotide-associated protein-2 (CNAP-2) cDNA.
XX
KW Cyclic nucleotide-associated protein-2; CNAP-2; human; cytostatic;
KW anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;
KW immunomodulatory; anti-asthmatic; anti-anemic; anti-diabetic; diagnosis;
KW anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;
KW anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;
KW anti-infectivity; anti-allergic; vasotropic; immunosuppressive;
KW hypotensive; gene therapy; prevention; treatment; arteriosclerosis;
KW cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;
KW neurological; vision; reproductive; smooth muscle; ss.
XX
OS Homo sapiens.


```

FH Key Location/Qualifiers
FT CDS 31..4014
FT /tag= a
FT /product= "Human CNAP-2 protein"
FT /note= "Shares 24% identity to Aquifex pyrophilus
FT esterase 28LC"
FT sig_peptide 31..132
FT /tag= b
FT mat_peptide 133..4011
FT /tag= c
FT /product= "Mature CNAP-2 protein"
FT misc_binding 136..165
FT /tag= d
FT /note_moiety= "Primer or Probe"
FT /note= "Useful for amplification or hybridisation
FT techniques"
XX
XX
PN W0200014248-A1.
XX
PN 16-MAR-2000.
XX
XX
PF 03-SEP-1999; 99WO-US20287.
XX
PR 04-SEP-1998; 98US-0148904.
XX
XX (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;
XX
DR WPI; 2000-256994/22.
DR P-PSDR; Y70474.
XX
XX
PT Isolated cyclic nucleotide associated proteins useful for preventing,
PT diagnosing and treating cell proliferative, autoimmune/inflammatory,
PT neurological, vision, reproductive and smooth muscle disorders -
XX
XX Claim 9; Page 70-71; 78pp; English.
XX
CC The present sequence is the cDNA encoding human cyclic nucleotide
CC associated protein-2 (CNAP-2), identified in Incyte clone 3149674,
CC that is isolated from ADREN004 CDNA library. It is expressed in
CC nervous, reproductive, cardiovascular and haematopoietic/immune tissues.
CC CNAP sequences may be used for prevention, treatment and diagnosis of
CC diseases associated with altered CNAP expression such as, cell
CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia,
CC lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/
CC inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple
CC sclerosis and psoriasis), neurological disorders (e.g. epilepsy,
CC Alzheimer's/ Parkinson's disease and strokes), vision disorders (e.g.
CC conjunctivitis, glaucoma, cataracts and retinitis pigmentosa),
CC reproductive disorders (e.g. infertility, uterine fibroids, ectopic
CC pregnancies and impotence) and smooth muscle disorders (e.g. angina,
CC anaphylactic shock, Kearns-Sayre syndrome and hypertension). It can also
CC be used for gene therapy.
XX
XX Sequence 4228 BP; 793 A; 1365 C; 1324 G; 746 T; 0 other;

```


Search completed: January 30, 2001, 21:29:29
Job time: 19255 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:11:29 ; Search time 64.84 Seconds
(without alignments)
155.458 Million cell updates/sec

Title: US-08-799-910-10_COPY_71_156

Perfect score: 442

Sequence: 1 VRRQLPVEPNPAKRLFL...EPSDYALDLSTFLOQHPAAF 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPTREMBL_15:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	442	100.0	193	4	075353
2	84.5	19.1	637	4	094878
3	81.5	18.4	102	5	0972D3
4	81.5	18.4	452	5	03UAS5
5	81.5	18.4	468	5	09XY65
6	80.5	18.2	302	12	09QM69
7	78	17.6	4957	4	014687
8	78	17.6	5262	4	014686
9	77	17.4	1568	4	095785
10	77	17.4	2424	5	09VZ48
11	76	17.2	148	2	09JZ87
12	76	17.2	801	5	09A3K6
13	76	17.2	1677	5	000805
14	74	16.7	114	12	09YX08
15	74	16.7	955	11	088287
16	74	16.7	1040	5	09VH04
17	74	16.7	1561	11	088286
18	73.5	16.6	925	12	069138
19	73	16.5	1029	12	09YMX0

20	72	16.3	559	5	09VQ23
21	72	16.3	852	4	09NP71
22	71	16.1	359	5	09TYW7
23	71	16.1	1882	3	09P898
24	70.5	16.0	244	12	069126
25	70.5	16.0	1900	4	015022
26	70.5	16.0	2605	4	09UK25
27	70.5	16.0	2715	4	09UMN6
28	70	15.8	148	2	09JUV2
29	70	15.8	1233	2	09RI32
30	70	15.8	2971	4	09Y5L9
31	69.5	15.7	277	4	09NXN6
32	69.5	15.7	562	4	09NVJ4
33	69.5	15.7	745	4	09P2N8
34	69.5	15.7	757	6	09N015
35	69.5	15.7	831	4	09UMS6
36	69.5	15.7	1887	3	012533
37	69	15.6	418	5	018503
38	69	15.6	646	4	09ULX6
39	69	15.6	646	4	09UGM0
40	69	15.6	1454	5	010463
41	69	15.6	2157	11	09Z1R1
42	69	15.6	2187	11	P70670
43	68.5	15.5	701	4	09NXX8
44	68.5	15.5	734	4	09NSE6
45	68.5	15.5	821	4	09UII8

ALIGNMENTS

RESULT 1
075353 PRELIMINARY; PRT; 193 AA.
AC 075353;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE ANTI-DEATH PROTEIN.
CN IEX-1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98369175; PubMed=9703517;
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival."
RL Science 281:998-1001(1998).
DR EMBL; AF039067; AAC32558.1;
DR EMBL; AF071596; AAC72344.1;
SQ SEQUENCE 193 AA; 21028 MW; 7927D9D3FFBC7C57 CRC64;

Query Match 100.0%; Score 442; DR 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 8.7e-40;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VRRQLPVEPNPAKRLFL...EPSDYALDLSTFLOQHPAAF 86
Db 108 VRRQLPVEPNPAKRLFL...EPSDYALDLSTFLOQHPAAF 193
Qy 61 PFNLTSEPSDYALDLSTFLOQHPAAF 86
Db 168 PFNLTSEPSDYALDLSTFLOQHPAAF 193

RESULT 2
094878 PRELIMINARY; PRT; 637 AA.
ID 094878
AC 094878;

DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence-update)
 DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
 DE KIAA0781 PROTEIN (FRAGMENT).
 GN KIAA0781.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euleleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 5:277-286(1998).
 DR EMBL; AB018324; BAA34501.1;
 DR INTERPRO; IPR000449;
 DR INTERPRO; IPR002965;
 DR PRINTS; PRO1217; PRICHEXTNSN.
 FT NON_TER
 FT 1
 SQ SEQUENCE 637 AA; 70516 MW; DEBFB005BA9B9CDD CRC64;

Query Match 19.1%; Score 84.5; DB 4; Length 637;
 Best Local Similarity 33.3%; Pred. NO. 0.41;
 Matches 30; Conservative 12; Mismatches 27; Indels 21; Gaps 5;
 QY 7 VEENPAKRLLEL---LTIIVFCOILMAEEGVP-----APLPEDAPNAASLAP----- 52
 DB 423 LGHEHRLQRLFLQKQSQVQAFVQMQIAESSYPOPSQQLPLPROETPPPSQOAPPPSILT 482
 QY 53 TPVSPVLEPNLTSEPSDYALDLSTFLQOH 82
 DB 483 QPLSPVLEP---SSEOMQV---SPFLSOY 505

RESULT 3
 ID Q9T2D3 PRELIMINARY; PRT; 102 AA.
 AC Q9T2D3;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
 DE R160.5 PROTEIN.
 GN R160.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifen L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
 RA Therry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;

RA Nelson J., Gattung S.;
 RT "The sequence of C. elegans cosmid R160.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF099001; AAC68734.1;
 SQ SEQUENCE 102 AA; 11426 MW; AFD366C3932EB26C CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 102;
 Best Local Similarity 24.7%; Pred. NO. 0.13;
 Matches 19; Conservative 19; Mismatches 28; Indels 11; Gaps 3;
 QY 11 NPAKRLLEL---LTIIVFCOILMAEEGVPAPLPEDAPNAASLAPVPSPVLEPFN-- 63
 DB 7 NPAIQFAFVVILVCAVCICMLIERAGFCPLPPELRIRIVTIP-PTPTIAAPVNR 65
 QY 64 ---LTSEPSDYALDLST 77
 DB 66 QVVIKIDPTESDVMDS 82

RESULT 4
 ID Q9UAS5 PRELIMINARY; PRT; 452 AA.
 AC Q9UAS5;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
 DE DEVELOPMENTAL-SPECIFIC PROTEIN CONZAB (FRAGMENT).
 GN CONZAB.
 OS Euplotes crassus.
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;
 OC Euplotes.
 OX NCBI_TaxID=5936;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ST9, ST11;
 RX MEDLINE=97206572; PubMed=9172827;
 RA Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;
 RT "Conjugation-specific genes in the ciliate Euplotes crassus: gene
 RT expression from the old macronucleus."
 RL J. Eukaryot. Microbiol. 44:1-11(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ST9, ST11;
 RA Jacobs M.E., Ling Z., Klobutcher L.A.;
 RT "conzab encodes a novel and abundant protein targeted to the
 RT developing macronucleus in Euplotes crassus."
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF116196; AAD29625.1;
 FT NON_TER
 FT 1
 SQ SEQUENCE 452 AA; 49763 MW; A5C13C49ECA998EA CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 452;
 Best Local Similarity 34.5%; Pred. NO. 0.61;
 Matches 30; Conservative 10; Mismatches 30; Indels 17; Gaps 5;
 QY 3 RQLVPEEP----NPAKRLFLLLLTIVFCOILMAEEGVPAPLPEDAPNAASLAP----P 54
 DB 316 RQSPTPLPILLENPAKEIASLSKMI---QETAIESIPTQPP-----VSSIQPSAPVHP 367
 QY 55 VSPVLEPFN-LTSEPSDYALDLSTFLQ 80
 DB 368 SAPSPSSTLASEPRDIIIPDPSAALK 394

RESULT 5
 Q9XY65

ID	QXKY65	PRELIMINARY;	PRT;	468	AA.
AC	Q9XY65;				
DT	01-NOV-1999	(TrEMBLrel. 12, Created)			
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)			
DT	01-MAY-2000	(TrEMBLrel. 13, Last annotation update)			
GN	CONZAB	PROTEIN.			
GN	CONZAB				
OS	Euplotes crassus.				
OC	Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;				
OC	Euplotes.				
OX	NCBI_TaxID=5936;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=G1;				
RC	MEDLINE=97206572; PubMed=9172827;				
RX	Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;				
RA	"Conjugation-specific genes in the ciliate Euplotes crassus: gene				
RT	expression from the old macronucleus.";				
RL	J. Eukaryot. Microbiol. 44:1-11(1997).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=G1;				
RC	MEDLINE=98391748; PubMed=9722644;				
RX	Klobutcher L.A., Gygas S.E., Podoloff J.D., Vermeesch J.R.,				
RA	Price C.M., Tebeau C.M., Jahn C.L.;				
RT	"Conserved DNA sequences adjacent to chromosome fragmentation and				
RL	telomere addition sites in Euplotes crassus.";				
RL	Nucleic Acids Res. 26:4230-4240(1998).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=G1;				
RC	Ling Z., Klobutcher L.A.;				
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.				
RL	EMBL; AF061334; AAD31724.1;				
DR	SEQUENCE	468	AA;	51457	MW: 70996125317C592A
DR					CRC64;

Query Match 18.4%; Score. 81.5; DB 5; Length 468;
 Best Local Similarity 34.5%; Pred. NO. 0.63;
 Matches 30; Conservative 10; Mismatches 30; Indels 17; Gaps 5;

QY 3 RQLPVEEP---NPAKRLJLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPT-----P 54
||| | |||| : | :: | : | | | : | : |
Db 332 ROSPTPLDILENPAKEIASLSKMI--QETAIESITTPQPQ-----VSSIQPSAPVHP 383

QY 55 VSPVLEPFN-LTSEPSDYALDI,STFLQ 80
: | | | : | | | | | | :
Db 384 SAPSLEPSSTLASEPRDIIPDP,SAALK 410

RESULT	6	
Q9QM69	PRELIMINARY;	PRT: 302 AA.
ID	Q9QM69	
AC	Q9QM69;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)	
DE	33 KDA PROTEIN.	
OS	fowl adenovirus 8.	
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.	
OX	NCBI_TaxID=66295;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Ojkic B., Nagy E.;	
RT	"The DNA sequence of fowl adenovirus 8.";	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: AF083975; AAD50348.1; -.	
SQ	SEQUENCE 302 AA; 34763 MW; BF34DF42FF026CCC CRC64;	

Query Match 18.2%; Score 80.5; DB 12; Length 302;
Best Local Similarity 26.4%; Pred. No. 0.51;
Matches 24; Conservative 26; Mismatches 30; Indels 11; Gaps 4;

[illegible]

RESULT 7

```

014687
ID 014687 PRELIMINARY; PRT: 4957 AA.
AC 014687;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE ALR.
DE ALR.
DE ALR.
OS Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN SEQUENCE FROM N.A.
RA MEDLINE=97388474; PubMed=9247308;
RX Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canani E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALU-1 involved in acute leukemia and to Drosophila
RT trithorax";
RL Oncogene 15:549-560(1997).
RL EMBL: AF010404; AAC51735.1; -.
DR INTERPRO: IPR001214; -.
DR INTERPRO: IPR001822; -.
DR INTERPRO: IPR001955; -.
DR INTERPRO: IPR002955; -.
DR PFAM: PF00628; PHD; 3.
DR PFAM: PF00856; SET; 1.
DR PRINTS: PR01217; PRICEXTENS.
DR PROSITE: PS00398; RECOMBINASES_2; UNKNOWN_1.
SQ SEQUENCE 4957 AA; 531848 MW; 10265632E1419C8D CRC64

```

Query Match 17.6%; Score 78; DB 4; Length 4957;
Best Local Similarity 30.4%; Pred. NO. 17;
Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

```
QY      1 VRRLPVEEPNPAKRLLELLITVFCQILMAEEGVAPLPEDAPNAASLAPTVPSPVLE 60
       :::: | || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     4299 LKOESPADEPPQHRYTYNVSNLDVRQL-----SAPPDEPSPPPSPAPSPASPTE 4351
```

Qy 61 PF-NLTSEP 68
Db 4352 PLVELPTEP 4360

RESULT

014686	PRELIMINARY;	PRT;	5262 AA.
ID	014686		
AC	014686;		
DT	01-JAN-1998 ('TREMBlrel. 05, Created)		
DT	01-JAN-1998 ('TREMBlrel. 05, Last sequence update)		
DT	01-OCT-2000 ('TREMBlrel. 15, Last annotation update)		
DE	ALR.		
GN	ALR.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
	{1}		
RN	SEQUENCE FROM N.A.		
RP			

RX MEDLINE=97388474; PubMed=9247308;
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
 RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
 RA Canaan E.,
 RT "Structure and expression pattern of human ALR, a novel gene with
 RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
 RL trithorax";
 RL Oncogene 15:549-560(1997).
 DR EMBL: AF010403; AAC51734.1; -
 DR INTERPRO: IPR001214; -
 DR INTERPRO: IPR001822; -
 DR INTERPRO: IPR001841; -
 DR INTERPRO: IPR001965; -
 DR INTERPRO: IPR002965; -
 DR PFAM: PF00628; PHD: 5;
 DR PFAM: PF00856; SET: 1;
 DR PRINTS: PR01217; PRICHEXTENS.
 DR PROSITE: PS00398; RECOMBINASES_2; UNKNOWN_1.
 SQ SEQUENCE 5262 AA; 564181 MW; 26B7C74CAD17B44 CRC64;

Query Match 17.6%; Score 78; DB 4; Length 5262;
 Best Local Similarity 30.4%; Pred No. 18;
 Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

OY 1 VRQLVEEDNPAKRLFLLLTIVFCQILMAEGVPAPLPDPDAPNAASLAPTPVSPVLE 60
 DB 4604 LQAESPAPEPTQHRYTYNVSNLDVRQL-----SAPPEEPSPPPLAPSPASPTE 4656
 QY 61 PF-NLTSEP 68
 DB 4657 PLVELDPT 4655

RESULT 9
 ID O95785 PRELIMINARY; PRT: 1568 AA.
 AC O95785;
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DE HUMAN HOMOLOG OF MUS MUSCULUS WIZL PROTEIN (HUMAN HOMOLOG OF MUS
 DE MUSCULUS WIZS PROTEIN) (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
 RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Barnes J.,
 RA Dangnanan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,
 RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
 RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RA Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of a 3.5 Mb region in 19p13.1 between OLFR and
 RT JUND";
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

[2]
 RN SEQUENCE OF 783-1568 FROM N.A.
 RP Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
 RA Burkhardt-Schultz K., Gordon L., Dias J., Brower A., Ramirez M.,
 RA Stilwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,
 RA Barnes J., Dangnanan L., Erlar A., Christensen M., Georgescu A.,
 RA Avila J., Liu S., Attix C., Andreise T., Trankheim M.,
 RA Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,
 RA Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D.,
 RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of an ~1.5 Mb contig in 19p13.1 between OLFR and
 RT D19S885";
 RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC007059; AAD19818.1; -
 DR EMBL: AC006128; AAC97985.1; -
 DR EMBL: AC007059; AAD19817.1; -
 DR INTERPRO: IPR000822; -
 DR PFAM: PF00096; ZF-C2H2; 10.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2; 8.
 KW Zinc-finger; Metal-binding; DNA-binding.
 FT NON_TER 1
 SQ SEQUENCE 1568 AA; 170077 MW; 74F0D34D45F565D0 CRC64;

Query Match 17.4%; Score 77; DB 4; Length 1568;
 Best Local Similarity 40.4%; Pred No. 6.6;
 Matches 19; Conservative 3; Mismatches 17; Indels 8; Gaps 1;

OY 30 MAEGVPAPLPDPDAPNAASLAPTPVSPVLE-----PFNLITSEP 68
 DB 908 LALAGSPTPKPNPKSQLSUSPRASPAPKQPMQSEDEGGLNLTSGP 954

RESULT 10
 ID O9V248 PRELIMINARY; PRT: 2424 AA.
 AC O9V248;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DE 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE CG2174 PROTEIN.
 GN CG2174.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champé M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abayani A., An H.-J., Andrews-Bianknoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flooker A., Gong F., Garg N.S., Gelbart W.M., Glasser K.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Plutman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reiner B.C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue E., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003484; AAF47980.1;
DR HSSP: P06799; 1MND.
DR FLYBASE: FBgn0030252; CG2174.
DR INTERPRO: IPR000048;
DR INTERPRO: IPR000857;
DR INTERPRO: IPR001609;
DR PFAM: PF00063; myosin_head; 1.
DR PFAM: PF00612; IQ; 3.
DR PFAM: PF00784; MYTH4; 1.
DR PRINTS: PROU193; MYOSINHEAVY.
SQ SEQUENCE 2424 AA; 267616 MW; 8ADG2AA33F9AA5D0 CRC64;

Query Match 17.4%; Score 77; DB 5; Length 2424;
Best Local Similarity 36.7%; Pred. No. 10;
Matches 22; Conservative 7; Mismatches 29; Indels 2; Gaps 2;

QY 28 ILMAEGVPAPLPEDAPNAASLAPTPVSPLEPNLTSEPS-DYALDLSTFLOQHPAAF 86
Db 1950 LLPPAPSVAPPPPIRPPSPMAPPAPP-AQSPPTARSPPEPDYRTSSQVVKHVPAPF 2008

RESULT 11
ID Q9JZR7 PRELIMINARY; PRT: 148 AA.
AC Q9JZR7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ACYL COA THIOESTER HYDROLASE FAMILY PROTEIN.
GN NMB0925.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson M.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi C., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL: AE002444; AAF41333.1;
DR TIGR: NMB0925;
KW Hydrolase.
SQ SEQUENCE 148 AA; 16332 MW; 28E9B40D502D80B4 CRC64;

Query Match 17.2%; Score 76; DB 2; Length 148;
Best Local Similarity 40.5%; Pred. No. 0.75;
Matches 15; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 18 FLLLTIVFCQLMAEGVPAPLPEDAPNAASLAPTP 54
Db 112 YLVTEAVFTYVAIDAEGRNPRIKPGNPKLAGLLTP 148

RESULT 12
Q9W3K6 PRELIMINARY; PRT: 801 AA.
ID Q9W3K6
AC Q9W3K6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG2258 PROTEIN.
GN CG2258.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.H., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003442; AAF46320.1;
DR HSSP: P06241; 1SHF.
DR FLYBASE: FBgn0029997; CG2258.
DR INTERPRO: IPR001452;
DR PFAM: PF00018; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
SQ SEQUENCE 801 AA; 89233 MW; A0F5C824663DC391 CRC64;

Query Match 17.2%; Score 76; DB 5; Length 801;
Best Local Similarity 33.9%; Pred. No. 4.2;
Matches 21; Conservative 8; Mismatches 23; Indels 10; Gaps 2;

QY 26 CQILMAEGVPAPLPEDA--PNAASLAPTPVS-----PYLEPNLTSEPSDYALDL 75
Db 206 CTIOTFOOQOPLPAPPAPAPSPSAASATPTAASTSKAATGDDVAQNSITSEPTTTT 265

QY 76 ST 77
Db 266 TT 267

RESULT 13

```


000805
ID 000805 PRELIMINARY: PRT: 1677 AA.
AC 000805;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE P-GLYCOPROTEIN E.
OS Leishmania tropica.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LRC-L39;
RA Lafuente E., Castanys S., Camarro F.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U55381; AAB51191.1;
DR HSSP: P13569; INBD.
DR INTERPRO: IPR000504;
DR INTERPRO: IPR001140;
DR INTERPRO: IPR001617;
DR INTERPRO: IPR001865;
DR PFAM: PF00005; ABC_tran; 2.
DR PFAM: PF00664; ABC_membrane; 2.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00962; RIBOSOMAL_S2.1; UNKNOWN_1.
SQ SEQUENCE 1677 AA; 183006 MW; C49208921B71563A CRC64;

Query Match 17.2%; Score 76; DB 5; Length 1677;
Best Local Similarity 33.9%; Pred. No. 9;
Matches 20; Conservative 6; Mismatches 15; Indels 18; Gaps 2;

QY 37 APLP-----PEDANNAASLAPTPVSPVLEPN-----LTSEPSDYALDLST 77
||||| : : : : : ||| : : : : :
Db 698 APLPEAEAPLPDPQSSSAAPRPAGVTEPLNAGKSHGTGHGDAASSEPLSSSAQKST 756

RESULT 14
QYVQ8
ID QYVQ8 PRELIMINARY: PRT: 114 AA.
AC QYVQ8;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE LATE 33 K PROTEIN HOMOLOG.
OS fowl adenovirus 8.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=66295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC A-2A;
RX MEDLINE=98451338; PubMed=9780058;
RA Cao J.X., Kreil P.J., Nagy E.;
RT "Sequence and transcriptional analysis of terminal regions of the fowl
adenovirus type 8 genome."
RL J. Gen. Virol. 79:2507-2516(1998).
DR EMBL: AF021254; AAC71673.1;
SQ SEQUENCE 114 AA; 13266 MW; 866DEEC6636EABA6 CRC64;

Query Match 16.7%; Score 74; DB 12; Length 114;
Best Local Similarity 31.1%; Pred. No. 0.94;
Matches 19; Conservative 17; Mismatches 15; Indels 10; Gaps 3;

QY 32 ECGVPAPLP-----PEDANNAASLAPTPVSPVLEPN-----LTSEPSDYALDLST 83
||||| : : : : : ||| : : : : :
Db 12 EEPVPPPLPKRRRRANKKEPNASSETQLPEVSPAVS--DLKAEILNLLVEIESFVRKNP 69

QY 84 A 84
Db 70 S 70

RESULT 15
O88287
ID O88287 PRELIMINARY: PRT: 955 AA.
AC O88287;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE WIZ.
GN WIZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Matsumoto K., Ishii N., Yoshida S., Shiosaka S., Wanaka A.,
RA Tohyama M.;
RT "Molecular Cloning and Distinct Developmental Expression Pattern of
RT Spliced Forms of A Novel Zinc Finger Gene wiz in The Cerebellum.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB012266; BAA32791.1;
DR MGD: MGI:1332638; WIZ.
DR INTERPRO: IPR000822;
DR PFAM: PF00096; zf-C2H2; 5.
DR PRINTS: PR00048; ZINC_FINGER.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 5.
KW Zinc-finger; Metal-binding; DNA-binding.
SQ SEQUENCE 955 AA; 102777 MW; D4DD03A02BECCAAC CRC64;

Query Match 16.7%; Score 74; DB 11; Length 955;
Best Local Similarity 41.9%; Pred. No. 8.3;
Matches 18; Conservative 2; Mismatches 15; Indels 8; Gaps 1;

QY 34 GVPAPLPEDAPNNAASLAPTPVSPVLE-----PFNLTSSEP 68
||||| : : : : : ||| : : : : :
Db 300 GSPTPKNPEDKSPQLSPRLSPRAQCPQSEDEGLNLTSGP 342

Search completed: January 30, 2001, 23:11:30
Job time: 3995 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 21:29:34 ; Search time 57.93 seconds
(without alignments)
401.549 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPDVALDLSTFLOQHPAAF 156

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 912318 seqs, 149113690 residues

Total number of hits satisfying chosen parameters: 912318

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending_Patents_AA:*
- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
 - 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
 - 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
 - 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
 - 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
 - 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
 - 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
 - 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
 - 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
 - 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
 - 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
 - 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
 - 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
 - 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
 - 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
 - 21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
 - 22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
 - 23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
 - 24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
 - 25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
 - 26: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
 - 27: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
 - 28: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
1	823	100.0	156 11	US-08-799-910-10 Sequence 10, Appl
2	823	100.0	156 12	US-08-825-486-10 Sequence 10, Appl
3	823	100.0	156 12	US-08-826-248-10 Sequence 10, Appl
4	823	100.0	156 12	US-08-870-434-11 Sequence 11, Appl
5	823	100.0	156 13	US-08-925-588-10 Sequence 10, Appl

6	823	100.0	156	17	US-09-372-044-10	Sequence 10, Appl
7	794	96.5	151	22	US-60-197-873-17386	Sequence 17386, A
8	440	53.5	123	22	US-60-196-718-6102	Sequence 6102, Ap
9	425	51.6	83	1	PCT-US00-15136-168	Sequence 168, App
10	423	51.4	83	1	PCT-US00-15136-96	Sequence 96, Appl
11	422	51.3	83	1	PCT-US00-15136-167	Sequence 167, App
12	381	46.3	89	22	US-60-196-718-4945	Sequence 4945, Ap
13	358	43.5	71	1	PCT-US00-15136-169	Sequence 169, App
14	358	43.5	92	22	US-60-196-718-5439	Sequence 5439, Ap
15	340	41.3	72	1	PCT-US00-15136-170	Sequence 170, App
16	273.5	33.2	56	22	US-60-196-718-3936	Sequence 3936, Ap
17	114.5	13.9	1828	22	US-60-188-986-15	Sequence 15, Appl
18	114.5	13.9	2012	22	US-60-229-524-80	Sequence 80, Appl
19	114.5	13.9	2972	19	US-09-579-181-2	Sequence 2, Appl1
20	114.5	13.9	3041	28	US-60-258-275-409	Sequence 409, App
21	114.5	13.9	3118	19	US-09-579-181-1	Sequence 1, Appl1
22	112.5	13.7	649	22	US-60-191-637-28141	Sequence 28141, A
23	112.5	13.7	649	22	US-60-191-681-22750	Sequence 22750, A
24	112.5	13.7	1255	22	US-60-167-217-13778	Sequence 13778, A
25	112.5	13.7	1255	22	US-60-173-464-11201	Sequence 11201, A
26	112	13.6	747	18	US-09-417-507-31296	Sequence 31296, A
27	110	13.4	590	1	PCT-US99-22853B-2801	Sequence 2801, Ap
28	109	13.2	8991	11	US-08-714-741-32	Sequence 32, Appl
29	108	13.1	271	16	US-09-248-796-21050	Sequence 21050, A
30	108	13.1	271	22	US-60-096-409-21050	Sequence 21050, A
31	107.5	13.1	1013	22	US-60-191-637-39541	Sequence 39541, A
32	107.5	13.1	1013	22	US-60-191-681-30574	Sequence 30574, A
33	107	13.0	206	9	US-08-529-055-54	Sequence 54, Appl
34	106	12.9	1110	22	US-60-243-468-898	Sequence 898, App
35	105.5	12.8	1013	22	US-60-173-464-30071	Sequence 30071, A
36	105	12.8	296	22	US-60-173-464-18776	Sequence 18776, A
37	105	12.8	296	22	US-60-191-637-22907	Sequence 22907, A
38	105	12.8	296	22	US-60-191-681-18056	Sequence 18056, A
39	104	12.6	129	9	US-08-529-055-53	Sequence 53, Appl
40	102	12.4	331	22	US-60-173-464-29642	Sequence 29642, A
41	102	12.4	331	22	US-60-191-637-38878	Sequence 38878, A
42	102	12.4	331	22	US-60-191-681-30133	Sequence 30133, A
43	102	12.4	421	18	US-09-451-320-4244	Sequence 4244, Ap
44	101.5	12.3	193	9	US-08-529-055-49	Sequence 49, Appl
45	101.5	12.3	741	18	US-09-417-507-37506	Sequence 37506, A

ALIGNMENTS

RESULT 1
US-08-799-910-10
; Sequence 10, Application US/08799910
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/799,910
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996

ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-067-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-799-910-10

Query Match 100.0%; Score 823; DB 11; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRSCPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRSCPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASL 120
DB 61 RSRRLVPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOQHPAAF 156
DB 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOQHPAAF 156

RESULT 2
US-08-825-486-10
Sequence 10, Application US/08825486
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
TITLE OF INVENTION: CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,486
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-077-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-825-486-10

Query Match 100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRSCPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRSCPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASL 120
DB 61 RSRRLVPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOQHPAAF 156
DB 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOQHPAAF 156

RESULT 3
US-08-826-248-10
Sequence 10, Application US/08826248
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
TITLE OF INVENTION: CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,248
FILING DATE: 28-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011,787
FILING DATE: 16-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-079-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein

```
; FRAGMENT TYPE: internal
US-08-826-248-10

Query Match      100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156

RESULT 4
US-08-870-434-11
; Sequence 11, Application US/08870434
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870.434
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799.910
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-084
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; MOLECULE TYPE: protein
; US-08-870-434-11

Query Match      100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
```

```
; FRAGMENT TYPE: internal
US-08-925-588-10

Query Match      100.0%; Score 823; DB 13; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156

RESULT 5
US-08-925-588-10
; Sequence 10, Application US/08925588
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925.588
; FILING DATE: 08-Sep-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799.910
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-08-925-588-10

Query Match      100.0%; Score 823; DB 13; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156
```

RESULT 6
US-09-372-044-10
; Sequence 10, Application US/09372044A
; GENERAL INFORMATION:
; APPLICANT: Dean FALB et al.
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Treatment and Diagnosis of Cardiovascular Disease
; FILE REFERENCE: 7853-152
; CURRENT APPLICATION NUMBER: US/09/372.044A
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-372-044-10

Query Match 100.0%; Score 823; DB 17; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MCHSRSCPTMTILOAPTAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRSCPTMTILOAPTAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120
Db 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120
QY 121 APTVPSPVLEPNLTSEPSDYALDLSTFLQHPAAF 156
Db 121 APTVPSPVLEPNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 7
US-60-197-873-17386
; Sequence 17386, Application US/60197873
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81 US1 PRO
; CURRENT APPLICATION NUMBER: US/60/197,873
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 17386
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-197-873-17386

Query Match 96.5%; Score 794; DB 22; Length 151;
Best Local Similarity 100.0%; Pred. No. 7.5e-58;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MCHSRSCPTMTILOAPTAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRSCPTMTILOAPTAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120
Db 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120
QY 121 APTVPSPVLEPNLTSEPSDYALDLSTFLQ 151

Db 121 APTVPSPVLEPNLTSEPSDYALDLSTFLQ 151
RESULT 8
US-60-196-718-6102
; Sequence 6102, Application US/60196718
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: C1000456
; CURRENT APPLICATION NUMBER: US/60/196,718
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6102
; LENGTH: 123
; TYPE: PRT
; ORGANISM: HUMAN
US-60-196-718-6102

Query Match 53.5%; Score 440; DB 22; Length 123;
Best Local Similarity 93.5%; Pred. No. 7.4e-29;
Matches 86; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 65 VLYPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASLAPT 124
Db 32 LFLCLPLRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASLAPT 91
QY 125 VSPVLEPNLTSEPSDYALDLSTFLQHPAAF 156
Db 92 VSPVLEPNLTSEPSDYALDLSTFLQHPAAF 123

RESULT 9
PCT-US00-15136-168
; Sequence 168, Application PC/TUS0015136
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PS551PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 168
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-15136-168

Query Match 51.6%; Score 425; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 8.2e-28;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 70 VVRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASLAPTVPSPVL 129
Db 1 VVRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASLAPTVPSPVL 60
QY 130 EPNLTSEPSDYALDLSTFLQ 152
Db 61 EPNLTSEPSDYALDLSTFLQ 83

RESULT 10
PCT-US00-15136-96
; Sequence 96, Application PC/TUS0015136

GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PS551PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 96
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (83)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids

PCT-US00-15136-96

Query Match 51.4%; Score 423; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 1.2e-27;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 LPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAIPPPEDAPNAASLAPTPVSPVLEPFLN 134
Db 1 LPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAIPPPEDAPNAASLAPTPVSPVLEPFLN 60
QY 135 TSEPSDYALDLSTFLQOHPAAF 156
Db 61 TSEPSDYALDLSTFLQOHPAAF 82

RESULT 11
PCT-US00-15136-167
; Sequence 167, Application PC/TUS0015136
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PS551PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 167
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-15136-167

Query Match 51.3%; Score 422; DB 1; Length 83;
Best Local Similarity 98.8%; Pred. No. 1.4e-27;
Matches 82; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 70 VYRROLVPEEPNPAKRLFLLLTIVFCQILMAEEGVPAIPPPEDAPNAASLAPTPVSPVL 129
Db 1 LVYRROLVPEEPNPAKRLFLLLTIVFCQILMAEEGVPAIPPPEDAPNAASLAPTPVSPVL 60
QY 130 EPNLTSEPSDYALDLSTFLQOH 152
Db 61 EPNLTSEPSDYALDLSTFLQOH 83

RESULT 12
US-60-196-718-4945
; Sequence 4945, Application US/60196718
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000456
; CURRENT APPLICATION NUMBER: US/60/196,718
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4945
; LENGTH: 89
; TYPE: PRT
; ORGANISM: HUMAN
US-60-196-718-4945

Query Match 46.3%; Score 381; DB 22; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.7e-24;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRSCHTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGRHK 60
Db 1 MCHSRSCHTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGRHK 60
QY 61 RSRRLVLYPRV 70
Db 61 RSRRLVLYPRV 70

RESULT 13
PCT-US00-15136-169
; Sequence 169, Application PC/TUS0015136
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PS551PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 169
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-15136-169

Query Match 43.5%; Score 358; DB 1; Length 71;
Best Local Similarity 97.1%; Pred. No. 2.2e-22;
Matches 66; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCHSRSCHTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGRHK 60
Db 1 MCHSRSCHTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGRHK 60
QY 61 RSRRLVLYP 68
Db 61 RSRRLVLYP 68

RESULT 14
US-60-196-718-5439
; Sequence 5439, Application US/60196718
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000456
; CURRENT APPLICATION NUMBER: US/60/196,718
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5439
; LENGTH: 92
; TYPE: PRT
; ORGANISM: HUMAN
US-60-196-718-5439

Query Match 43.5%; Score 358; DB 22; Length 92;
Best Local Similarity 100.0%; Pred. No. 2.9e-22;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MCHSRSCPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRSCPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
QY 61 RSRRL 66
DB 61 RSRRL 66

RESULT 15
PCT-US00-15136-170
; Sequence 170, Application PC/TUS0015136
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: PS551PCT
; CURRENT APPLICATION NUMBER: PCT/US00/15136
; CURRENT FILING DATE: 2000-06-01
; EARLIER APPLICATION NUMBER: 60/138,629
; EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-15136-170

Query Match 41.3%; Score 340; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 6.8e-21;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MCHSRSCPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRSCPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
QY 61 RS 62
DB 61 RS 62

Search completed: January 30, 2001, 23:10:11
Job time: 6037 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 19:24:41 ; Search time 36.91 Seconds
(without alignments)
75.895 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPDVALDLSTFLQHPAAF 156

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	823	100.0	156	3	US-08-826-246-10
2	823	100.0	156	3	US-08-944-495-10
3	823	100.0	156	3	US-08-126-640-11
4	93.5	11.6	739	3	US-08-035-648-24
5	93.5	11.6	1596	3	US-08-356-952-3
6	93.5	11.4	905	2	US-08-574-959A-9
7	93.5	11.4	1135	2	US-08-574-959A-7
8	92	11.2	1400	1	US-08-080-255-7
9	92	11.2	1400	3	US-08-465-713-7
10	92	11.2	1400	4	PCT-US93-05857-7
11	90.5	11.0	214	1	US-08-217-327-4
12	88	10.7	468	2	US-08-390-000A-7
13	88	10.7	477	1	US-08-444-734A-4
14	88	10.7	477	1	US-08-087-772A-16
15	87	10.6	2414	1	US-08-227-536-2
16	87	10.6	2414	4	PCT-US95-04682-2
17	85.5	10.4	610	1	US-07-821-171B-6
18	85.5	10.4	610	1	US-08-119-262B-6
19	85.5	10.4	610	1	US-08-135-929A-11
20	85.5	10.4	610	1	US-08-234-265A-11
21	85.5	10.4	774	2	US-08-231-193A-42
22	85.5	10.4	774	2	US-08-486-273A-42
23	85.5	10.4	774	3	US-08-480-474-42
24	85.5	10.4	774	3	US-08-940-086A-42
25	85.5	10.4	1214	2	US-08-231-193A-54
26	85.5	10.4	1214	2	US-08-486-273A-54
27	85.5	10.4	1214	3	US-08-480-474-54
28	85.5	10.4	1214	3	US-08-940-086A-54

29	85.5	10.4	1219	2	US-08-231-193A-50
30	85.5	10.4	1219	2	US-08-486-273A-50
31	85.5	10.4	1219	3	US-08-480-474-50
32	85.5	10.4	1219	3	US-08-940-086A-50
33	85.5	10.4	1231	2	US-08-231-193A-48
34	85.5	10.4	1231	2	US-08-486-273A-48
35	85.5	10.4	1231	3	US-08-480-474-48
36	85.5	10.4	1231	3	US-08-940-086A-48
37	85.5	10.4	1236	2	US-08-231-193A-6
38	85.5	10.4	1236	2	US-08-486-273A-6
39	85.5	10.4	1236	3	US-08-480-474-6
40	85.5	10.4	1236	3	US-08-940-086A-6
41	85.5	10.4	1239	2	US-08-231-193A-52
42	85.5	10.4	1239	2	US-08-486-273A-52
43	85.5	10.4	1239	3	US-08-480-474-52
44	85.5	10.4	1239	3	US-08-940-086A-52
45	85.5	10.4	1244	2	US-08-231-193A-46

ALIGNMENTS

RESULT 1
US-08-826-246-10
; Sequence 10, Application US/08826246
; Patent No. 6048709
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,246
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-078-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-826-246-10

Sequence 50, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 52, Appl
Sequence 52, Appl
Sequence 52, Appl
Sequence 52, Appl
Sequence 46, Appl

Query Match 100.0%; Score 823; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 4e-75;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
Qy 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120
Db 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120
Qy 121 APTVPSPVLEPFNLTSPESDYALDLSTFLOQHPAAF 156
Db 121 APTVPSPVLEPFNLTSPESDYALDLSTFLOQHPAAF 156

RESULT 2

US-08-944-495-10
; Sequence 10, Application US/08944495
; Patent No. 6087477
; GENERAL INFORMATION:

; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,495
; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
US-08-944-495-10

Query Match 100.0%; Score 823; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 4e-75;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120
Db 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120
Qy 121 APTVPSPVLEPFNLTSPESDYALDLSTFLOQHPAAF 156
Db 121 APTVPSPVLEPFNLTSPESDYALDLSTFLOQHPAAF 156

RESULT 3

US-09-126-640-11
; Sequence 11, Application US/09126640A
; Patent No. 6099823
; GENERAL INFORMATION:

; APPLICANT: Falb, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/126,640A
; CURRENT FILING DATE: 1998-07-30

; EARLIER APPLICATION NUMBER: 08/870,434
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 08/799,910
; EARLIER FILING DATE: 1997-02-13

; EARLIER APPLICATION NUMBER: 60/011,787
; EARLIER FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 156

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-126-640-11

Query Match 100.0%; Score 823; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 4e-75;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120
Db 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120

Qy 121 APTVPSPVLEPFNLTSPESDYALDLSTFLOQHPAAF 156
Db 121 APTVPSPVLEPFNLTSPESDYALDLSTFLOQHPAAF 156

RESULT 4

US-09-035-648-24
; Sequence 24, Application US/09035648
; Patent No. 6100031
; GENERAL INFORMATION:

; APPLICANT: Shyjan, Andrew W.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL

; TITLE OF INVENTION: GROWTH AND PROLIFERATION

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/035,648
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/818,829
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meiklejohn, Ph.D., Anita L.
 ; REGISTRATION NUMBER: 35,283
 ; REFERENCE/DOCKET NUMBER: 07334/003001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-542-5070
 ; TELEFAX: 617-542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 24:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 739 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FRAGMENT TYPE: internal
 ; US-09-035-648-24

Query Match 11.6%; Score 95.5; DB 3; Length 739;
 Best Local Similarity 27.0%; Pred. No. 0.089;
 Matches 47; Conservative 14; Mismatches 60; Indels 53; Gaps 7;
 QY 16 APTAPSTIPGP-----RRSGPGEFTFDPLPEPAAAPAGRPS-----ASRGHR---- 59
 DB 264 APPYSAVTPPPDAFSGVSSPIAGPAQPPWPQAPWSPAFYDSSERIASRDERISVP 323
 QY 60 -KRS-----RRVLPVRRQLPVEEPNPAKRLFLLL----- 91
 DB 324 AKRTGILQAKRRSTTKPMFTKEPKVSPN--ELLSSLQNSEGKRGTCGAGDGSPEEDY 381
 QY 92 -----TIVFCOILMAEGVAPLPEDAPNAASLAP-TPVSPVLEPFLNLTSEP 138
 DB 382 LSLGAEACNFMOSSAKQKTPPPVAPKPAVKSSSQPVTPVSPVWSPGVNPTQP 435

RESULT 5
 US-09-356-952-3
 ; Sequence 3, Application US/09356952
 ; Patent No. 6117663
 ; GENERAL INFORMATION:
 ; APPLICANT: Boriack-Sjodin, Ann
 ; APPLICANT: Margarit, S. M.
 ; APPLICANT: Bor-Sogli, Dafna
 ; APPLICANT: Cole, Philip
 ; APPLICANT: Kurlyan, John
 ; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: 600-1-228N
 ; CURRENT APPLICATION NUMBER: US/09/356,952
 ; EARLIER FILING DATE: 1999-07-19
 ; EARLIER APPLICATION NUMBER: 60/093,631
 ; EARLIER FILING DATE: 1998-07-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 1596
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; US-09-356-952-3

Query Match 11.6%; Score 95.5; DB 3; Length 1596;
 Best Local Similarity 28.3%; Pred. No. 0.22;
 Matches 52; Conservative 19; Mismatches 68; Indels 45; Gaps 13;

QY 4 SRSCHPTWYTLQATTPAPSTIPGHRG--SQPEITFTDPLPEPAAAPAGRPSASKGHKR 61
 DB 1354 TESCADMAQKRAAP-DAP-TLP-PRDGELSP-----PIPPRLNIHSTGISYLRSHGKS 1404
 QY 62 SRRV-----LYPR---VVRQLPVE-----BPNPAKRLFLLLTIVFCOILMAEGSV 105
 DB 1405 KEFVGNSSLLLPNTSSIMIRNSALEKRAAATSQPNQAAGPISTWTLVTVSOAVATDEPL 1464
 QY 106 PAPLPEDAPNAASLAPT--PVSPVLEPF--NLTSEP-----SDYALDLSTFLQO----H 152
 DB 1465 PLPISP-----AASSSTTTSLTPAMSPMSPNIPSHPVVESTSSSVNQLRMROQOQOOTH 1519
 QY 153 PAAP 156
 DB 1520 PAIY 1523

RESULT 6
 US-08-574-959A-9
 ; Sequence 9, Application US/08574959A
 ; Patent No. 5962224
 ; GENERAL INFORMATION:
 ; APPLICANT: Jaekyoon Shin, Insil Joong, Rauna K. Vadiamudi
 ; APPLICANT: and Jack L. Strominger
 ; TITLE OF INVENTION: P62 POLYPEPTIDES, RELATED POLYPEPTIDES
 ; TITLE OF INVENTION: AND USES THEREFOR
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, Suite 510
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/574,959A
 ; FILING DATE: 19-DEC-95
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mandragouras, Amy E.
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: DFN-008
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)227-5941
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 905 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-574-959A-9

Query Match 11.4%; Score 93.5; DB 2; Length 905;
 Best Local Similarity 23.1%; Pred. No. 0.18;
 Matches 45; Conservative 17; Mismatches 62; Indels 71; Gaps 8;
 QY 8 HPTMTILQ-----APTAPSTIPGPRGSGPE-----LFTFDPLPEPAAAPAGRPSAS 55
 DB 400 HFRVPLQPMGPTCTPAPVPLLRPHRPSGPHRSILRAPCPQAWAPCPQAPCPGAPMPS 459
 QY 56 RG-----HKKRSRR--VLYP----- 68
 DB 460 AGPVSEPWTSTTANLLGLLSRPSVCPPLLPENHNRAGSNEDPILAPSGTPTPTIPD 519
 QY 69 -----RVVRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVAPLPEDAPNAASLAPT 123
 DB 520 ETFGGRVPRPAFVHVHDKESASOVSLESDSDSVVIVPEGLP-PLPP---PPPSGATPP 575


```
;
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465.713
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/080.255
; FILING DATE: 17 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-05857-7
;
;
; Query Match 11.2%; Score 92; DB 3; Length 1400;
; Best Local Similarity 22.1%; Pred. No. 0.42;
; Matches 31; Conservative 17; Mismatches 40; Indels 52; Gaps 5;
;
QY 15 QATPAPSTIPGRRGSGPEITFDPLPE-----PAAAPAGRPSASRGRKKS 62
DB 465 QKPTPSAREDPAPKKSS-----EPPPKPVEEKSEGCNVSAPGCSKQATTPASRKS 518
;
QY 63 RVLVPRVV-----RRQLPVEEPNPAKRLLFLLTIVFCOILMAEEGVAPLPP 111
DB 519 KOVSOPALVIPQPTTGPGRKEVPKTTTPEPKK-----KQPPPP 558
;
QY 112 EDAPNAA---SLAPTPVSPV 128
DB 559 ESGPEQSKQKKVAPRPSIPV 578
;
RESULT 10
PCT-US93-05857-7
; Sequence 7, Application PC/TUS9305857
; GENERAL INFORMATION:
; APPLICANT: Board of Regents
; APPLICANT: The University of Texas System
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; OPERATING SYSTEM: PC-DOS/MS-DOS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05857
; FILING DATE: 19930617
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: 07/900.689
; FILING DATE: 17/06/92
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-05857-7
;
;
; Query Match 11.2%; Score 92; DB 4; Length 1400;
; Best Local Similarity 22.1%; Pred. No. 0.42;
; Matches 31; Conservative 17; Mismatches 40; Indels 52; Gaps 5;
;
QY 15 QATPAPSTIPGRRGSGPEITFDPLPE-----PAAAPAGRPSASRGRKKS 62
DB 465 QKPTPSAREDPAPKKSS-----EPPPKPVEEKSEGCNVSAPGCSKQATTPASRKS 518
;
QY 63 RVLVPRVV-----RRQLPVEEPNPAKRLLFLLTIVFCOILMAEEGVAPLPP 111
DB 519 KOVSOPALVIPQPTTGPGRKEVPKTTTPEPKK-----KQPPPP 558
;
QY 112 EDAPNAA---SLAPTPVSPV 128
DB 559 ESGPEQSKQKKVAPRPSIPV 578
;
RESULT 11
US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E
; APPLICANT: Barton, Kenneth A
; TITLE OF INVENTION: Immobilized proteins in Cotton Fiber
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,327
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-217-327-4

Query Match 11.0% Score 90.5; DB 1; Length 214;
Best Local Similarity 27.5%; Pred. No. 0.066;
Matches 38; Conservative 8; Mismatches 59; Indels 33; Gaps 5;
Qy 9 PTMTILOQTPAPSTIPGRRGSGEIFTDP-----LPEPAAPAGRPASRCHRRKS 62
Db 54 PTATPPPVSTPTTSSPPPTVTPSPVSTPPSPPPATPPASPPPTATPPASPPPTATP 113
Qy 63 RRVLYPRVRLQPLVEENPAPRLLELLLTIVFCILMAEGVPAPLPPEADAPNAASLAP 122
Db 114 PPASPPATP---PPATPPA-----TPPATPPAPLASPPAT 149
Qy 123 TP-VSPVLEPFLNLTSEPS 139
Db 150 VPALSPVQTP--LTSPPA 165

RESULT 12
US-08-390-000A-7
Sequence 7, Application US/08390000A
Patent No. 5985363
GENERAL INFORMATION:
APPLICANT: Sealion, Stuart C.
TITLE OF INVENTION: Cloning and Expression of
TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,000A
FILING DATE: 17-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6923-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-390-000A-7

Query Match 10.7% Score 88; DB 2; Length 468;
Best Local Similarity 24.2%; Pred. No. 0.29;
Matches 50; Conservative 19; Mismatches 44; Indels 94; Gaps 12;
Qy 17 PTPAPSTIPGRRGSGEIFTDPPLPEPAAPAGRPAS--RGHRRSRRLVYPRVRRQ 74
Db 264 PPSPPSPVPAPAPPICP-----PRPAAAAATAPLANGRAGRPPSLV----ALREQ 311

Qy 75 -----LPVEEPNPAKH-----LFLLT-----JVCO- 97
Db 312 KAKTLGIIMGVFTLCWLPPFLANVKAHRELVPDLFVFNLMGYANSFNIYCRS 371
Qy 98 --ILMAEGV-----PAPLPPEADAPNAAS-----LA 121
Db 372 PDFRKAQGLCCARRARRRHATHGDRPRASGCLARPGP-PP--SYCAASUDDUDDUUVG 428
Qy 122 PTPVSPVLEP-----NLTSEPSDYALD 144
Db 429 ATPPARLLEPWAGCNGGAADSDSLD 455
RESULT 13
US-08-444-734A-4
Sequence 4, Application US/08444734A
Patent No. 5610282
GENERAL INFORMATION:
APPLICANT: Sibley, David R.
APPLICANT: Monsma, Frederick J.
APPLICANT: Mahan, Lawrence C.
APPLICANT: McVittie, Loris D.
TITLE OF INVENTION: cDNA encoding the rat D1 dopamine
TITLE OF INVENTION: receptor linked to adenylyl cyclase activation and
TITLE OF INVENTION: expression of the receptor protein in plasmid-transfect
TITLE OF INVENTION: cell lines
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson and Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,734A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/029,917
FILING DATE: 03-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/548,714
FILING DATE: 06-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: NIH065.001FWJ
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-444-734A-4

Query Match 10.7% Score 88; DB 1; Length 477;
Best Local Similarity 24.2%; Pred. No. 0.3;
Matches 50; Conservative 19; Mismatches 44; Indels 94; Gaps 12;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 18:47:31 ; Search time 54.97 Seconds
(without alignments)
97.039 Million cell updates/sec

Title: US-08-799-910-10
Perfect score: 823
Sequence: 1 MCHSRCHPTMTILQAPTA.....EPSDYALDLSTFLQHPAAF 156

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_36:**
1: /cgn2_2/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	823	100.0	156	21	Y45017
2	816	99.1	156	18	W36006
3	107	13.0	206	18	W14574
4	104	12.6	129	18	W14573
5	101.5	12.3	190	18	W14569
6	101	12.3	180	18	W14562
7	101	12.3	183	18	W14570
8	100	12.2	783	19	W37151
9	100	12.2	787	19	W37152
10	100	12.2	802	19	W37153
11	98.5	12.0	168	21	Y68732
12	96.5	11.7	179	21	Y68731

13	95.5	11.6	739	18	W25790
14	95.5	11.6	1596	21	Y68821
15	94.5	11.5	520	16	R78185
16	94	11.4	1290	20	Y06809
17	94	11.4	1290	21	Y32216
18	94	11.4	1291	20	Y01687
19	93.5	11.4	185	18	W14566
20	93.5	11.4	905	18	W31186
21	93.5	11.4	1135	18	W31185
22	92.5	11.2	572	18	W31855
23	92.5	11.2	763	18	W31852
24	92	11.2	166	18	W14558
25	92	11.2	167	18	W14558
26	92	11.2	1400	15	R44514
27	92	11.2	1400	17	R92705
28	92	11.2	3910	14	R38470
29	92	11.2	3910	16	R66462
30	91	11.1	1012	20	Y17406
31	90.5	11.0	214	17	R86913
32	90	10.9	171	20	Y29034
33	90	10.9	171	20	Y29074
34	90	10.9	254	20	Y29061
35	89.5	10.9	666	19	W72911
36	89.5	10.9	666	20	Y21928
37	89	10.8	788	21	Y54466
38	89	10.8	807	21	Y54467
39	89	10.8	1291	20	Y01689
40	89	10.8	1291	20	Y06810
41	89	10.8	3969	15	R52971
42	88.5	10.8	187	18	W14579
43	88.5	10.8	188	18	W14580
44	88.5	10.8	688	21	Y57898
45	88.5	10.8	711	19	W80602

ALIGNMENTS

RESULT 1
Y45017
ID Y45017 standard; Protein; 156 AA.
AC Y45017;
XX
XX
DT 31-MAY-2000 (first entry)
XX
DE Protein encoded by fchd605 gene.

XX fchd605 gene; human; cardiovascular disease; oncogenic disorder;
XX diabetic retinopathy; fibroproliferative disorder; atherosclerosis;
XX TGF-beta signalling pathway; TGF; Transforming growth factor;
XX pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;
XX vascularisation; cytostatic; antidiabetic; ophthalmological.

OS Homo sapiens.

PN WO200006206-A1.

PD 10-FEB-2000.

XX 30-JUL-1999; 99WO-US17394.

XX 30-JUL-1998; 98US-0126640.

XX (MILL-) MILLENNium PHARM INC.

PI Falb DA;

DR WPI: 2000-205414/18.

DR N-PSDB: 250711.

PT Identifying substances for ameliorating symptoms of fibroproliferative diseases or oncogenic related disorders.

XX

PS Example; Fig 5; 214pp; English.

XX

CC The patent discloses methods for the treatment and diagnosis of cardiovascular diseases by novel human genes which are differentially expressed in different cardiovascular disease states. Compositions which can modify MCP-beta signalling pathway are identified by screening. CC These are used therapeutically to treat fibroproliferative and oncogenic disorders, especially TGF (transforming growth factor)-beta related disorders, including diabetic retinopathy, arteriosclerosis, pancreatic cancer, angiogenesis, inflammation, fibrosis, tumour growth and vascularisation. The present sequence is the protein product of fchd605 gene which is up-regulated in monocytes treated with oxidised LDL (low density lipoprotein) can be used to design cardiovascular disease treatment strategies. Depending on whether the up-regulation has a pathogenic or protective effect treatment methods can be designed to increase or decrease the activity of the protein product of the gene.

XX

SQ Sequence 156 AA;

Query Match 100.0%; Score 823; DB 21; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.4e-69;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEITFDPLPEPAAAPACRPSASRGHRK 60
DB 1 mchsrchpmtllqaptapstipgprgsgpeitfdplpepaaapagrpasrghrk 60
OY 61 RSRRLVYPRVVRQIPVEEPNPAKRLLFLLITIVFCQILMAEGVPAPLPEDAPNAASL 120
DB 61 rsrtrlvprvvrqipveepnpakrllfllitvfcqilmaegvpaplpdpapnaasl 120
OY 121 APTPVSPVLEPFNLTSEPSDYALDLSTFLQHPAAF 156
DB 121 aptpvsplvlepfntsepsdyaldlstflqhpaf 156

RESULT 2

W36006

ID W36006 standard; Protein; 156 AA.

XX AC W36006;

XX DT 03-MAR-1998 (first entry)

XX DE Human Fchd605 gene product.

XX KW Fchd605 gene; differential expression; monocyte; human;
KW foam cell; cardiovascular disease; arteriosclerosis; ischaemia;
KW reperfusion; hypertension; restenosis; arterial inflammation;
KW therapy; diagnosis; drug screening; marker.

XX OS Homo sapiens.

XX PN WO9730065-A1.

XX PD 21-AUG-1997.

XX PF 14-FEB-1997; 97WO-US02291.

XX PR 13-FEB-1997; 97US-0799910.

XX PR 16-FEB-1996; 96US-0011787.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PF Falb DA;

XX DR WPI: 1997-424966/39.

XX DR N-PSDB; T94471.

XX New genes differentially expressed in cardiovascular disease - used
for diagnosis, drug screening and treatment of cardiovascular

PT disease, e.g. atherosclerosis, restenosis, hypertension, etc

XX Example 6; Fig 5; 163pp; English.

XX This protein is encoded by the novel human fchd605 gene (see T94471) that is up-regulated in monocytes treated with oxidised low density lipoproteins that simulate the conditions under which foam cells develop during atherogenesis. The protein has sequence similarity to the mouse gly96 gene and to EST T49532. Novel fchd531, fchd540, fchd545, fchd602 and fchd605 genes (see T94467-71) provide a fingerprint for the study of cardiovascular diseases, including atherosclerosis, ischaemia/reperfusion, hypertension, restenosis and arterial inflammation. Methods are provided for the diagnosis, monitoring in clinical trials, screening for therapeutically effective compounds, and treatment of cardiovascular diseases based on discoveries regarding the expression patterns of these novel genes.

XX Sequence 156 AA;

Query Match 99.1%; Score 816; DB 18; Length 156;
Best Local Similarity 99.4%; Pred. No. 3.3e-68;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEITFDPLPEPAAAPACRPSASRGHRK 60
DB 1 mchsrchpmtllqaptapstipgprgsgpeitfdplpepaaapagrpasrghrk 60
OY 61 RSRRLVYPRVVRQIPVEEPNPAKRLLFLLITIVFCQILMAEGVPAPLPEDAPNAASL 120
DB 61 rsrtrlvprvvrqipveepnpakrllfllitvfcqilmaegvpaplpdpapnaasl 120
OY 121 APTPVSPVLEPFNLTSEPSDYALDLSTFLQHPAAF 156
DB 121 aptpvsplvlepfntsepsdyaldlstflqhpaf 156

RESULT 3

W14574

ID W14574 standard; Protein; 206 AA.

XX AC W14574;

XX DT 28-OCT-1997 (first entry)

XX DE Streptococcus pneumoniae PspA central region.

XX KW PspA; pneumococcal surface protein; vaccine; otitis media;
KW meningitis; bacteraemia; pneumonia.

XX OS Streptococcus pneumoniae strain Dbl5.

XX FH Key Location/Qualifiers

XX FT Misc-difference 50

XX FT /note= "unidentified amino acid"

XX PN WO9709994-A1.

XX PD 20-MAR-1997.

XX PF 16-SEP-1996; 96WO-US14819.

XX PR 15-SEP-1995; 95US-0529055.

XX PA (UABR-) UAB RES FOUND.

XX PI Briles DE, Brooks-Walter A, Grain MJ, Hollingshead S;
XX PI McDaniel LS, Swiatlo E, Tart R, Yother J;

XX DR WPI: 1997-202002/18.

XX Streptococcus pneumoniae surface protein PspC and truncated PspA -

SQ XX CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC

Qy Db Oy

RESULT 7

PS XX CC CC CC

This sequence shows the central portion, including the C-terminus of the alpha-helix region and some of the proline-rich region, of pneumococcal surface protein A (PspA) of *Streptococcus pneumoniae* strain 0922134c. Comparison of the N-terminal and central regions (W1453-57 and W14562-91) of PspA polypeptides from different pneumococcal strains can be used to divide the strains into several families based on sequence homologies. PspA polypeptides, or fragments of them, can be used in vaccines to protect animals against *S. pneumoniae* infection and hence for the prevention of diseases such as otitis media, meningitis, bacteraemia and pneumonia. The sequence of the 3' half of the PspA alpha-helical region and the immediate 5' tip of the coding sequence are likely to be the critical sequences for predicting PspA cross-reactions and vaccine composition.

Query Match	12.3%;	Score 101;	DB 18;	Length 180;
Best Local Similarity	27.4%;	Pred. No. 0.032;		
Matches	31;	Conservative	8;	Mismatches 34;
				Indels 40;
				Cans

RESULT 7

AC W14570;

Streptococcus pneumoniae PspA central region.

Streptococcus pneumoniae strain Bq9739.

PD 20-MAR-1997.

15-SEP-1995; 95US-0529055.

Briles DE. Brooks-Wal

WPI; 1997-202002/18.

Example 6; Fig 13; 296pp; English.

This sequence shows the central portion, including the C-terminus of the alpha-helix region and some of the proline-rich region, of pneumococcal surface protein A (pspA) of *Streptococcus pneumoniae*

CC This protein comprises novel murine neural Mena+. Its amino acid
 CC sequence was deduced from a cDNA clone (see V02998) obtained from
 CC

XX
PS
XX
Example 4; Page 60-63; 77pp; English.

CC This protein comprises novel murine neural Mena+. Its amino acid
 CC sequence was deduced from a cDNA clone obtained from a mouse brain
 CC cDNA library. Two other isoforms, neural Mena+ (see W37151) and
 CC neural Mena++ (see W37153), are also disclosed. Unlike mammalian
 CC Ena (Mena, see W37148), neural Mena isoforms exhibit neural
 CC tissue-specific distribution. Based on the disclosed Mena and Evi
 CC genes (see also V02996-98) and proteins (see also W37148-49), a
 CC variety of methods and compositions are provided for screening,
 CC isolating and characterising endogenous and exogenous factors,
 CC drugs and therapeutic agents useful to evaluate and/or control
 CC cytoskeletal dynamic events involved in normal and abnormal cell
 CC morphology, adhesion, motility, growth and/or differentiation. A
 CC method of detecting a modulator of Mena activity/expression is
 CC claimed.

XX Sequence 787 AA;

Query Match 12.2%; Score 100; DB 19; Length 787;

Best Local Similarity 26.0%; Pred. No. 0.21; Mismatches 51; Indels 46; Gaps 6;

Matches 38; Conservative 11;

Oy 2 CHRSCHPTMT-ILQAPTPAPSTIPGPRGSGPEIFDPLPFAAAP-----ACRPSASR 56

Db 456 cgsqaspppgtplastpskpsvlpssaga-----pasaetplnplgdssase 505

Oy 57 GHRKRSRVLYPRVVRQLPVEEPNPAKRLFLILLITIVFCQLMAEGCVPAPLPPDAPN 116

Db 506 -----pglqaasqpaesptpglglv-----gppapppppplps 538

Oy 117 ----AASLAPTPVSPVLEPNLTSEP 138

Db 539 gpayasalpppppppppppplstgpp 564

RESULT 10

ID W37153 standard; Protein; 802 AA.

XX W37153;

DT 06-JUL-1998 (first entry)

DE Mouse neural Mena+++ protein.

KW Neural Mena+++ protein; mammalian Ena; Enabled protein; Evi protein;
 KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
 KW cell growth; cell motility; mouse.

OS Mus musculus.

XX W09801755-A1.

XX 15-JAN-1998.

XX 03-JUL-1997; 97WO-US11669.

XX 05-JUL-1996; 96US-0675815.

PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX Gertler FB, Niebuhr K, Soriano P, Wehland J;

XX WPI; 1998-101197/09.

XX Detection of modulators of Mena and Ena-VASP-like genes and proteins
 PT - used in control of cytoskeletal dynamic events in normal and
 PT abnormal cell morphology, adhesion, motility, growth and
 PT differentiation

XX Example 4; Page 63-65; 77pp; English.

XX

CC This protein comprises novel murine neural Mena+. Its amino acid
 CC sequence was deduced from a cDNA clone obtained from a mouse brain
 CC cDNA library. Two other isoforms, neural Mena+ (see W37151) and
 CC neural Mena++ (see W37153), are also disclosed. Unlike mammalian
 CC Ena (Mena, see W37148), neural Mena isoforms exhibit neural
 CC tissue-specific distribution. Based on the disclosed Mena and Evi
 CC genes (see also V02996-98) and proteins (see also W37148-49), a
 CC variety of methods and compositions are provided for screening,
 CC isolating and characterising endogenous and exogenous factors,
 CC drugs and therapeutic agents useful to evaluate and/or control
 CC cytoskeletal dynamic events involved in normal and abnormal cell
 CC morphology, adhesion, motility, growth and/or differentiation. A
 CC method of detecting a modulator of Mena activity/expression is
 CC claimed.

XX Sequence 802 AA;

Query Match 12.2%; Score 100; DB 19; Length 802;

Best Local Similarity 26.0%; Pred. No. 0.21;

Matches 38; Conservative 11; Mismatches 51; Indels 46; Gaps 6;

Oy 2 CHRSCHPTMT-ILQAPTPAPSTIPGPRGSGPEIFDPLPFAAAP-----ACRPSASR 56

Db 471 cgsqaspppgtplastpskpsvlpssaga-----pasaetplnplgdssase 520

Oy 57 GHRKRSRVLYPRVVRQLPVEEPNPAKRLFLILLITIVFCQLMAEGCVPAPLPPDAPN 116

Db 521 -----pglqaasqpaesptpglglv-----gppapppppplps 553

Oy 117 ----AASLAPTPVSPVLEPNLTSEP 138

Db 554 gpayasalpppppppppppplstgpp 579

RESULT 11

ID Y68732 standard; Protein; 168 AA.

XX Y68732;

XX 05-MAY-2000 (first entry)

XX Amino acid sequence of high mobility group (HMG)-Y/R protein.

KW Retinoic acid receptor; RAR; high mobility group protein;
 KW HMG protein; nuclear receptor; antiproliferative; anti-inflammatory;
 KW antipsoriasis; anticancer.

OS Homo sapiens.

XX W0200001717-A2.

XX 13-JAN-2000.

XX 28-JUN-1999; 99WO-US14801.

XX 01-JUL-1998; 98US-0108298.

XX (ALLR) ALLERGAN SALES INC.

XX Nagpal S, Chandraratna RA, Ghosh C;

XX WPI; 2000-170999/15.

XX Identifying modulators of nuclear receptors, useful e.g. as anticancer
 PT and anti-inflammatory agents, according to their effect on binding
 PT between the receptor and high-mobility group protein -

XX Claim 28; Fig 1B; 43pp; English.

XX The present sequence represents a high mobility group (HMG)-Y/R protein.
 CC HMG sequences that interact with retinoic acid receptors (RARs) in a

CC Ligand dependent manner to induce gene expression can be identified
 CC using the method of the invention. The specification describes a
 CC method for identifying compounds that modulate a nuclear receptor.
 CC The method comprises incubating a polypeptide containing a
 CC ligand-binding region of nuclear receptor with test compound in
 CC presence of second polypeptide containing the nuclear receptor-binding
 CC part of a HMG protein, and detecting any alteration in binding between
 CC the two polypeptides. The method is useful for identifying potential
 CC antiproliferative, anti-inflammatory, antipsoriasis and anticancer
 CC agents.

XX
 SQ Sequence 168 AA;

Query Match 12.0%; Score 98.5; DB 21; Length 168;

Best Local Similarity 27.7%; Pred. No. 0.05;

Matches 39; Conservative 9; Mismatches 52; Indels 41; Gaps 4;

OY 17 PTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGPSASGHRKSRRLVYPRVVRQL- 75

DB 33 ppkpevptpkr-----prgpkyskknwrrkrkrasrrsprrrss 74

OY 76 -PVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPPEPAPNAASLAPTPVSPVLEPENL 134

DB 75 dpcvpapwhrsslglldsf-----aplp-----pppqpqahhhrl 113

OY 135 PSEPSDYALDSTFTLQHPAA 155

DB 114 wpppsstcaltttlhstpaa 134

RESULT 12

ID Y68731
 AC Y68731 standard; Protein; 179 AA.

XX
 XX Y68731;

DT 05-MAY-2000 (first entry)

DE Amino acid sequence of high mobility group (HMG)-R protein.

XX Retinoic acid receptor; RAR; high mobility group protein;

KW HMG protein; nuclear receptor; antiproliferative; anti-inflammatory;

KW antipsoriasis; anticancer.

XX
 XX Homo sapiens.

OS
 XX WO200001717-A2.

PN
 XX 13-JAN-2000.

PD
 XX 28-JUN-1999; 99WO-US14801.

PF
 XX 01-JUL-1998; 98US-0108298.

PR
 XX (ALLR) ALLERGAN SALES INC.

XX
 XX Nagpal S, Chandraratna RA, Ghosh C;

XX
 XX WPI: 2000-170999/15.

DR
 XX Identifying modulators of nuclear receptors, useful e.g. as anticancer

PT
 XX and anti-inflammatory agents, according to their effect on binding

PT
 XX between the receptor and high-mobility group protein

XX
 PS Claim 24; Fig 1B; 43pp; English.

XX
 CC The present sequence represents a high mobility group (HMG)-R protein.

CC HMG sequences that interact with retinoic acid receptors (RARs) in a

CC ligand dependent manner to induce gene expression can be identified

CC using the method of the invention. The specification describes a

CC method for identifying compounds that modulate a nuclear receptor.

CC
 CC The method comprises incubating a polypeptide containing a

CC ligand-binding region of nuclear receptor with test compound in
 CC presence of second polypeptide containing the nuclear receptor-binding
 CC part of a HMG protein, and detecting any alteration in binding between
 CC the two polypeptides. The method is useful for identifying potential
 CC antiproliferative, anti-inflammatory, antipsoriasis and anticancer
 CC agents.

XX
 SQ Sequence 179 AA;

Query Match 11.7%; Score 96.5; DB 21; Length 179;

Best Local Similarity 27.4%; Pred. No. 0.083;

Matches 40; Conservative 9; Mismatches 56; Indels 41; Gaps 4;

OY 12 TILQAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGPSASGHRKSRRLVYPRVVRQL- 71

DB 39 talvsgqkpevptpkr-----prgpkyskknwrrkrkrasrrspr 80

OY 72 RRLQ--PVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPPEPAPNAASLAPTPVSPVL 129

DB 81 rrrssdpcvpapwhrsslglldsf-----aplp-----pppqpqah 119

OY 130 EFPNLTSEPSDYALDSTFTLQHPAA 155

DB 120 hhhrlwpppsstcaltttlhstpaa 145

RESULT 13

ID W25790 standard; Protein; 739 AA.

XX
 XX W25790;

XX
 XX 11-MAR-1998 (first entry)

DE Gene 036 product differentially expressed in colon tumour cells.

XX Colon tumour; colon cancer; differential expression; gene 036;

KW human; diagnosis; gene therapy; tumour suppressor.

XX
 XX Homo sapiens.

OS
 XX WO9733551-A2.

PN
 XX 18-SEP-1997.

PD
 XX 14-MAR-1997; 97WO-US04191.

PF
 XX 15-MAR-1996; 96US-0013438.

PR
 XX (MILL-) MILLENNIUM PHARM.

PA
 XX Shyjan AW;

XX
 XX WPI: 1997-470615/43.

DR
 XX N-PSDE; T91708.

XX
 XX Gene 036 with reduced or amplified expression in tumour cells -

PT
 XX used, optionally with genes 097, 030 and 036 or their protein

PT
 XX products, for diagnosis and treatment of colonic cancer

XX
 PS Claim 38; Fig 2; 117pp; English.

XX
 CC This protein is encoded by gene 036 (see T91708), which was

CC identified by differential display analysis as being expressed at

CC a higher level in normal colon tissues than in cancerous colon

CC tissues. Gene 036 is a candidate tumour suppressor gene. A

CC correlation was found between an increase in the expression level

CC of gene 036 and a decrease in a colon cell's tumour potential

CC Hence, methods that increase the level of expression of gene 036

CC may inhibit or slow the progression to tumours and cancer, e.g.

CC colon cancer. The 036 protein is used in claimed methods for

CC treating a patient suffering from a disorder associated with

CC insufficient expression of gene 036 protein, and for identifying
CC compounds that modulate 036 protein activity. Such compounds are
CC useful in the diagnosis, prevention and treatment of tumours and
CC cancers.

XX
SQ Sequence 739 AA;

Query Match 11.6%; Score 95.5; DB 18; Length 739;
Best Local Similarity 27.0%; Pred. NO. 0.5;
Matches 47; Conservative 14; Mismatches 60; Indels 53; Gaps 7;
QY 16 APTAPSTIPGP-----RRGSGPEIFTFDPLPEPAAPAGRPDS-----ASRGHR----- 59
DB 264 appysavtppdpafgrvssplagpdpqpwpqapwsqpfydsseriesdrisvvp 323
QY 60 -RRS-----RRVLYPRVRRQLPVEEFPNPAKRLFLLL----- 91
DB 324 akrtgilqearsttkpmftfkpkvsnp--ellsllqnsegkrgtgaggdsgpeedy 381
QY 92 -----TIVFCQIILMAEGVPAPLPPEDAPNAASIAP-TPVSPVLEPNLTSEP 138
DB 382 lslgaacnfmqssakqtkppvpakpvaksssgvptpvpwpsvpgvaptqp 435

RESULT 14
Y68821
ID Y68821 standard; protein; 1596 AA.

XX Y68821;
AC XX
DT 16-MAY-2000 (first entry)
XX
XX Amino acid sequence of a Drosophila Son of sevenless (Sos) protein.
DE Ras; Son of sevenless; Sos; crystal; Ras-Sos complex; cancer;
KW protein coordinate data.
XX Drosophila melanogaster.
OS W0200005258-A1.
XX
XX
XX 03-FEB-2000.
XX
XX 20-JUL-1999; 99WO-US16348.
XX
XX 21-JUL-1998; 98US-0119794.
XX (UYRQ) UNIV ROCKEFELLER.
XX
XX Horiack-Sjodin A, Margarit SM, Bar-Sagi D, Cole P, Kuriyan J;
XX WPI; 2000-182647/16.
XX
XX
XX Novel crystals comprising a Ras-Son of sevenless complex, useful for
XX screening drugs useful in cancer treatment

XX Disclosure: Page 169-175; 224pp; English.
XX
XX The specification describes a crystal complex comprising at least a
XX Sos contacting region of a Ras protein and at least a Son of sevenless
XX (Sos) protein catalytic region fragment, that effectively diffracts
XX X-rays. Ras and Sos form a tight complex. Sos does not impede the
XX binding sites for the nucleotide base and the ribose of GTP or GDP
XX and thus the Ras-Sos complex maintains a structure that permits
XX nucleotide release and rebinding. The crystals are used for the
XX determination of the atomic coordinates of the complex to a resolution
XX of more than 5.0 Angstrom. The crystals, or a dataset comprising the
XX three-dimensional coordinates obtained from the crystals, is useful
XX for identifying an agent that stabilizes the Ras-Sos complex. The
XX crystals are also useful for identifying agents that inhibit the
XX formation of Ras-Sos complex. Ras and Sos fragments are useful for
XX growing a crystal of a protein-ligand complex. Agents that stabilize

CC or inhibit the formation of Ras-Sos complex are useful in the
CC treatment of cancer. The present sequence represents a Drosophila
CC Sos protein.
XX
SQ Sequence 1596 AA;

Query Match 11.6%; Score 95.5; DB 21; Length 1596;
Best Local Similarity 28.3%; Pred. No. 1.2;
Matches 52; Conservative 19; Mismatches 68; Indels 45; Gaps 13;
QY 4 SRSCHPTMTILQAPTPAPSTIPPRG--SGPEIFTFDPLPEPAAPAGRPASRGRHKR 61
DB 1354 tescadmqrqap-dap-tlp-prdgelspp-----pipprlnhstgisylrqshgks 1404
QY 62 SRRV-----LYPR-----VVRQLPVE-----EPNPAKRLLLFLLLIVFCQIILMAEGV 105
DB 1405 kefvgnssllipntssimifrnssalekraatsqpnqaaagpiettitvtvsqavatdepl 1464
QY 106 PAPLPEDAPNAASLAPT--PVSPVLEPF--NLTSEP-----SDYALDLSTFLQO----H 152
DB 1465 plisp-----aassttsltpbampspnpsphvestsssyahqlrmrqgqqqth 1519
QY 153 PAAF 156
DB 1520 paiv 1523

RESULT 15
R78185
ID R78185 standard; Protein; 520 AA.
XX R78185;
AC XX
DT 09-FEB-1996 (first entry)
XX
XX Protein sequence of PEA3-beta -an ETS transcription factor.
DE
DE Transcription factor; probe; reverse transcription; PCR; primer;
KW expression vector; E.coli; COS cell; ras; cancer cell multiplication;
KW polyoma virus; transformation.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Misc-difference 84 /note= "encoded by CTC"
FT Misc-difference 126 /note= "encoded by AAC"
FT Misc-difference 500 /note= "encoded by ACC"
FT
XX JP07145197-A.
XX
XX 06-JUN-1995.
XX
XX 25-NOV-1993; 93JP-0295393.
XX
XX 25-NOV-1993; 93JP-0295393.
XX (EISA) EISAI CO LTD.
XX (HIRA/) HIRANO T.
XX WPI; 1995-237197/31.
XX N-P5DB; Q91769.
XX
XX ETS transcription factor activated by ras - may be used in the study
XX of cancer cell proliferation and the proliferation of the polyoma
XX virus
XX Claim 1; Page 7-9; 9pp; Japanese.
XX
XX The amino acid sequence of the novel ETS transcription factor family

CC member - PEA3-beta. The gene was isolated from a HepG2 cell line cDNA
 CC library. The probe for the screening was prepared by reverse
 CC transcription on HepG2 mRNA followed by PCR using primers Q91770-1,
 CC to produce a probe of 170-200 bp. The screening isolated the full
 CC length sequence of the transcription factor. The gene was inserted into
 CC the expression vectors pBluescript KS and pCDV1 for expression of the
 CC protein in E.coli and COS7 cells, respectively. The ETS transcription
 CC factor has specificity for and is activated by ras. It is useful as a
 CC reagent in studies for the elucidation of the mechanism of cancer cell
 CC multiplication or polyoma virus transformation of cells.

XX Sequence 520 AA;

Query Match 11.5%; Score 94.5; DB 16; Length 520;
 Best Local Similarity 22.4%; Pred. No. 0.42;
 Matches 38; Conservative 13; Mismatches 62; Indels 57; Gaps 7;
 QY 3 HRSCHPTMTILQAPTPAPSTIPGRRGSGPEITFDPLPEPAAAGRPASRGHRKRS 62
 Db 162 hapaagpvqgv--gpapaphslpep-----gpqqqtfa-vprpqpqlqmp-----kmm 207
 QY 63 RRVLYPRVVRQQLPVEEPNPAKRLFLLLTIVFCOILMAERGVP----- 106
 Db 208 pengypsegrfqrlsep-----chfpqpqpgvpgdnrpsyhrqmsepiv 252
 QY 107 --APLPPED-----APNAASLAPTPVSPVLEPENTSEPSDYALD 144
 Db 253 paappppggfkqeyhdpkyehgvqgmpgppahgfgspmgikqepdrdcvd 302

Search completed: January 30, 2001, 22:05:37
 Job time: 11886 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 18:47:25 ; Search time 1450.12 Seconds
(without alignments)
1246.746 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468
Perfect score: 258
Sequence: 1 GTCGGCGCCACGTCGCCAGT.....AGCAACACCGCGCGCCTTC 258

Scoring table: IDENTITY_NUC %
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estrol1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estrol0:*
84: em_estroll1:*
85: em_estroll2:*
86: em_estroll3:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16: *
118: em_estp17: *
119: em_estp18: *
120: em_estp19: *
121: em_estp20: *
122: em_estp21: *
123: em_estp22: *
124: em_estp23: *
125: em_estp24: *
126: em_estp25: *
127: em_estp26: *
128: em_estp27: *
129: em_estp28: *
130: em_estp29: *
131: em_estp30: *
132: em_estp31: *
133: em_estp32: *
134: em_estp33: *
135: em_estp34: *
136: em_estp35: *
137: em_estp36: *
138: em_estp37: *
139: em_estp38: *
140: em_estp39: *
141: em_estp40: *
142: em_estp41: *
143: em_estp42: *
144: em_estp43: *
145: em_estp44: *
146: em_estp45: *
147: em_estp46: *
148: em_estp47: *
149: em_estp48: *
150: em_estp49: *
151: em_estp50: *
152: em_estp51: *
153: em_estp52: *
154: em_estp53: *
155: em_estp54: *
156: em_estp55: *
157: em_estp56: *
158: em_estp57: *
159: em_estp58: *
160: em_estp59: *
161: em_estp60: *
162: em_estp61: *
163: em_estp62: *
164: em_estp63: *
165: em_estp64: *
166: em_estp65: *
167: em_estp66: *
168: em_estp67: *
169: em_estp68: *
170: em_estp69: *
171: em_estp70: *
172: em_estp71: *
173: em_estp72: *
174: em_estp73: *
175: em_estp74: *
176: em_estp75: *
177: em_estp76: *
178: em_estp77: *
179: em_estp78: *
180: em_estp79: *
181: em_estp80: *
182: em_estp81: *
183: em_estp82: *
184: em_estp83: *
185: em_estp84: *
186: em_estp85: *
187: em_estp86: *
188: em_estp87: *
189: em_estp88: *

190: gb_gss25: *
191: gb_gss26: *
192: gb_gss27: *
193: gb_gss28: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	258	100.0	411	25	A1824906	A1824906 wb03e06.x
2	258	100.0	544	107	BE384949	BE384949 601276496
3	258	100.0	951	107	BE383865	BE383865 601273192
4	256.4	99.4	765	106	R1275966	R1275966 601120727
5	254.4	98.6	762	110	BE621346	BE621346 601493736
6	249.4	96.7	916	137	BE882130	BE882130 601505227
7	245.4	95.1	603	106	BE273386	BE273386 601143493
8	244.6	94.8	463	7	AA410666	AA410666 z130q09.r
9	242	93.8	585	97	AW960290	AW960290 EST372361
10	236.8	91.8	810	109	HE563592	HE563592 601334706
11	223	86.4	447	147	W52268	W52268 zc46a10.r1
12	217.2	84.2	513	11	AA743819	AA743819 ob01d04.s
13	205.4	79.6	212	142	H80073	H80073 yv80g11.r1
14	203	78.7	1070	137	BE878459	BE878459 601487866
15	197.8	76.7	504	2	AA143404	AA143404 zo66e09.r
16	195.6	75.8	502	90	AW465331	AW465331 BP230018B
17	195.6	75.8	523	135	BE751344	BE751344 203178 MA
18	195.6	75.8	547	135	BE750116	BE750116 201294 MA
19	195.6	75.8	557	139	BF040186	BF040186 BP250005A
20	195.6	75.8	571	139	BF045952	BF045952 BP250003B
21	195.6	75.8	594	37	AV617905	AV617905 AV617905
22	192	74.4	346	38	AV683186	AV683186 AV683186
23	189.2	73.3	545	139	BF043065	BF043065 BP250026B
24	188.8	73.2	617	38	AV690660	AV690660 AV690660
25	181	70.2	184	143	N57203	N57203 yw91h09.r1
26	173.6	67.3	701	135	BE735733	BE735733 601304556
27	166	64.3	389	146	W39753	W39753 zc80b07.r1
28	163.8	63.5	174	137	BE939840	BE939840 RC6-UT001
29	160.8	62.3	597	37	AV594384	AV594384 AV594384
30	155	60.1	443	147	W47587	W47587 zc35b02.r1
31	155	60.1	806	139	BF025858	BF025858 601670012
32	154.4	59.8	628	96	AW916683	AW916683 EST347987
33	151.2	58.6	467	19	A1326484	A1326484 mq23c12.y
34	146.4	56.7	182	6	AA356940	AA356940 EST65572.y
35	145.8	56.5	150	89	AW367127	AW367127 MR0-HT015
36	144.8	56.1	487	139	BF023106	BF023106 ux04d04.y
37	143.2	55.5	471	109	BE554106	BE554106 ux38c03.y
38	139.4	54.0	1155	135	BE786915	BE786915 601477783
39	139.2	54.0	385	147	W77628	W77628 me68a09.r1
40	139.2	54.0	460	94	AW740625	AW740625 ur02h11.y
41	137	53.1	353	142	N45391	N45391 yw97d08.r1
42	136	52.7	425	2	AA122977	AA122977 mq23c12.r
43	127.6	49.5	443	147	W71619	W71619 me40d11.r1
44	126	48.8	459	26	AI876358	AI876358 uk74a05.y
45	125.4	48.6	476	12	AA833402	AA833402 ub58b04.r

ALIGNMENTS

RESULT 1
A1824906/C
LOCUS A1824906 411 bp mRNA EST 16-DEC-1999
DEFINITION wb03e06.x1 NCI-CCAF-GC6 Homo sapiens cDNA clone IMAGE:2304610 3' similar to SW:1EX1.HUMAN P46695 RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 ; mRNA sequence.
ACCESSION A1824906
VERSION A1824906.1 GI:5445577
KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 411)
 AUTHORS NCJ-CGAP <http://www.ncbi.nlm.nih.gov/ncjgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCJ-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
 Insert Length: 557 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 341.

FEATURES

source
 1..411
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2304610"
 /clone_lib="NCJ_CGAP_CC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Plasmid DNA from the normalized library
 NCJ-CGAP.GC4 was prepared, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from a pool of 5,000 clones made
 from the same library (cloneIDs 1257096-1258631,
 1469064-1470983, and 1475592-1476743). Subtraction by
 Bento Soares and M. Fatima Bonaldo."
 77 a 104 c 168 g 61 t 1 others

BASE COUNT
 ORIGIN

Query Match 100.0%; Score 258; DB 25; Length 411;
 Best Local Similarity 100.0%; Pred. No. 1e-55;
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCGGGGCCAGTGCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
 DB 282 GTCGGGGCCAGTGCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 223
 QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGGCCCTGCCT 120
 DB 222 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGGCCCTGCCT 163
 QY 121 CCAGAGAGCGCCCTAACCGCCGATCCCTGGCGCCACCCCTGTGTGCCCGCTCTCGAG 180
 DB 162 CCAGAGAGCGCCCTAACCGCCGATCCCTGGCGCCACCCCTGTGTGCCCGCTCTCGAG 103
 QY 181 CCCTTTAACTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTTCTCCAG 240
 DB 102 CCCTTTAACTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTTCTCCAG 43
 QY 241 CAACACCGCGCCCTTC 258
 DB 42 CAACACCGCGCCCTTC 25

RESULT 2
 BE384949
 LOCUS

BE384949 544 bp mRNA EST 21-JUL-2000

DEFINITION

601276496F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3617543 5',
 mRNA sequence.
 BE384949
 VERSION BE384949.1 GI:9330314
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 544)
 AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC/DCTD/DTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM284 row: p column: 24
 High quality sequence stop: 540.

FEATURES

source
 1..544
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3617543"
 /clone_lib="NIH_MGC_20"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 98 a 225 c 131 g 90 t

BASE COUNT
 ORIGIN

Query Match 100.0%; Score 258; DB 107; Length 544;
 Best Local Similarity 100.0%; Pred. No. 1.1e-55;
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCGGGGCCAGTGCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
 DB 212 GTCGGGGCCAGTGCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 271
 QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGGCCCTGCCT 120
 DB 272 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGGCCCTGCCT 331
 QY 121 CCAGAGGAGCGCCCTAACCGCCGATCCCTGGCGCCACCCCTGTGTGCCCGCTCTCGAG 180
 DB 332 CCAGAGGAGCGCCCTAACCGCCGATCCCTGGCGCCACCCCTGTGTGCCCGCTCTCGAG 391
 QY 181 CCCTTTAACTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTTCTCCAG 240
 DB 392 CCCTTTAACTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTTCTCCAG 451
 QY 241 CAACACCGCGCCCTTC 258
 DB 452 CAACACCGCGCCCTTC 469

RESULT 3

BE383865
 LOCUS

BE383865 951 bp mRNA EST 21-JUL-2000
 DEFINITION 601273192F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3614255 5',

ACCESSION BE383865
 VERSION BE383865.1 GI:9329230
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 JOURNAL NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC/DCTD/DTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM276 row: g column: 24
 High quality sequence stop: 747.

FEATURES
 Source

1..951
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3614255"
 /clone_lib="NIH_MGC_20"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 210 a 324 c 253 g 164 t
 ORIGIN

Query Match 100.0%; Score 258; DB 107; Length 951;
 Best Local Similarity 100.0%; Pred. No. 1.le-55;
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCCGGCGCAGCTGCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
 DB 210 GTCCGGCGCAGCTGCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 269
 QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGGCTGAAGAGGGGTGTGCCGGCGCCCTGCC 120
 DB 270 CTCACCATCGTCTTCTGCCAGATCCCTGATGGCTGAAGAGGGGTGTGCCGGCGCCCTGCC 329
 QY 121 CCAGAGGAGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCCCTCTCCGAG 180
 DB 330 CCAGAGGAGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCCCTCTCCGAG 389
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCCTCCAG 240
 DB 390 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCCTCCAG 449
 QY 241 CAACACCGCGCCCTTC 258
 DB 450 CAACACCGCGCCCTTC 467

RESULT 4
 BE275966
 LOCUS HE275966 765 bp mRNA
 DEFINITION 601120727F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2967534 5',
 mRNA sequence.

ACCESSION BE275966
 VERSION BE275966.1 GI:9150928
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 JOURNAL NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC/DCTD/DTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM72 row: e column: 07
 High quality sequence start: 7
 High quality sequence stop: 699.

FEATURES
 Source

1..765
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2967534"
 /clone_lib="NIH_MGC_20"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 151 a 284 c 202 g 128 t
 ORIGIN

Query Match 99.4%; Score 256.4; DB 106; Length 765;
 Best Local Similarity 99.6%; Pred. No. 2.8e-55;
 Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTCCGGCGCAGCTGCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
 DB 257 GTCCGGCGCAGCTGCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 316
 QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGGCTGAAGAGGGGTGTGCCGGCGCCCTGCC 120
 DB 317 CTCACCATCGTCTTCTGCCAGATCCCTGATGGCTGAAGAGGGGTGTGCCGGCGCCCTGCC 376
 QY 121 CCAGAGGAGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCCCTCTCCGAG 180
 DB 377 CCAGAGGAGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCCCTCTCCGAG 436
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCCTCCAG 240
 DB 437 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCCTCCAG 496
 QY 241 CAACACCGCGCCCTTC 258
 DB 497 CAACACCGCGCCCTTC 514

RESULT 5
 BE275966
 LOCUS BE275966 762 bp mRNA
 DEFINITION 601493736F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3895736 5',
 mRNA sequence.

BE521346
 BE521346.1 GI:9892286
 EST.
 Homo sapiens
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 762)
 NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM674 row: h column: 09
 High quality sequence stop: 613.
 Location/Qualifiers
 1..762
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3895736"
 /clone_lib="NIH_MGC_70"
 /tissue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site 2: Sall; Cloned unidirectionally. Primer: Oligo df.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."
 BASE COUNT 155 a 288 c 193 g 126 t
 ORIGIN
 Query Match 98.6%; Score 254.4; DB 110; Length 762;
 Best Local Similarity 99.6%; Pred. No. 8.9e-55;
 Matches 255; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTCCGGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 209 GTCCGGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 268
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 CTCACATCGTTCGTGCGAGATCCTGATGGCTGAAGAGGTGTGCGGCGCCCTGCCT 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 269 CTCACATCGTTCGTGCGAGATCCTGATGGCTGAAGAGGTGTGCGGCGCCCTGCCT 328
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 121 CCAGAGACGCCCTAAGCGCGATCCCTGGCGGCCACCCCTGTCTCCCGCTCTCGAG 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 329 CCAGAGACGCCCTAAGCGCGATCCCTGGCGGCCACCCCTGTCTCCCGCTCTCGAG 388
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCTCCAG 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 389 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCTCCAG 448
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 241 CAACACCCGGCGCT 256
 |||||||||||||||
 Db 449 CAACACCCGGCGCTT 464
 |||||||||||||||
 RESULT 6
 BE882130
 LOCUS BE882130.1 GI:10330906
 DEFINITION 60150527P2 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906823 5',
 mRNA sequence.
 ACCESSION BE882130
 VERSION BE882130.1
 KEYWORDS EST.
 SOURCE human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 916)
 NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM703 row: f column: 08
 High quality sequence stop: 659.
 Location/Qualifiers
 1..916
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3906823"
 /clone_lib="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site 2: Sall; Cloned unidirectionally. Primer: Oligo df.
 Average insert size 2.1 kb.
 BASE COUNT 187 a 327 c 268 g 134 t
 ORIGIN
 Query Match 96.7%; Score 249.4; DB 137; Length 916;
 Best Local Similarity 99.6%; Pred. No. 1.7e-53;
 Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTCCGGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 180 GTCCGGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 239
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 CTCACATCGTTCGTGCGAGATCCTGATGGCTGAAGAGGTGTGCGGCGCCCTGCCT 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 240 CTCACATCGTTCGTGCGAGATCCTGATGGCTGAAGAGGTGTGCGGCGCCCTGCCT 299
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 121 CCAGAGACGCCCTAAGCGCGATCCCTGGCGGCCACCCCTGTCTCCCGCTCTCGAG 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 300 CCAGAGACGCCCTAAGCGCGATCCCTGGCGGCCACCCCTGTCTCCCGCTCTCGAG 359
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCTCCAG 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 360 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCTCCAG 419
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 241 CAACACCCGGC 251
 |||||||||||
 Db 420 CAACACCCGGC 430
 |||||||||||
 RESULT 7
 BE273386
 LOCUS BE273386.1
 DEFINITION 601143493F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:3507191 5',
 mRNA sequence.
 ACCESSION BE273386
 VERSION BE273386.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 603)
 NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM186 tow: j column: 24
High quality sequence stop: 579.

FEATURES source

1. .603
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3507191"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pORF7; Site: 1; XhoI; Site: 2; EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 112 a 241 c 154 g 96 t
ORIGIN

Query Match 95.1%; Score 245.4; DB 106; Length 603;
Best Local Similarity 99.2%; Pred. No. 1.7e-52;
Matches 257; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GTCCGGCGCCAGCTCCAGTCGAGAACCCAGCCAAAGGCTTCCTTCCTGCTG 60
DB 177 GTCCGGCGCCAGCTCCAGTCGAGAACCCAGCCAAAGGCTTCCTTCCTGCTG 236
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCCTGCCCT 120
DB 237 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCCTGCCCT 296
QY 121 CCAGAGAGCGCCCTACAGCGCGATCCTGTGGGCCCAACCCCTGTGTGCC-CCGTCCTCGA 179
DB 297 CCAGAGAGCGCCCTACAGCGCGATCCTGTGGGCCCAACCCCTGTGTGCC-CCGTCCTCGA 356
QY 180 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCTCGACCTTCCTCCA 239
DB 357 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCTCGACCTTCCTCCA 416
QY 240 GCAACACCCGCGCCCTTC 258
DB 417 GCAACACCCGCGCCCTTC 435

RESULT 8

AA410666 463 bp mRNA EST 18-MAY-1997
LOCUS zt30g09.r1 Soares ovary tumor NBH07 Homo sapiens cDNA clone
DEFINITION IMAGE:723904.5, similar to SW:GL96_MOUSE P46694 IMMEDIATE EARLY
PROTEIN GLY96. ;, mRNA sequence.

ACCESSION AA410666
VERSION AA410666.1 GI:2069789
KEYWORDS EST.
SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 463)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

TITLE JOURNAL COMMENT

Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, H., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R., and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL. Contact the IMAGE Consortium (info:image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 Et from Amerham
High quality sequence stop: 369.

FEATURES source

1. .463
/organism="Homo sapiens"
/db_xref="GDB:5935321"
/db_xref="taxon:9606"
/clone_lib="IMAGE:723904"
/clone="IMAGE:723904"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="organ: ovary; Vector: pMT3D (Pharmacia) with a modified polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGGTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pMT3 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 71 a 196 c 114 g 82 t
ORIGIN

Query Match 94.8%; Score 244.6; DB 7; Length 463;
Best Local Similarity 98.4%; Pred. No. 2.6e-52;
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTCCAGTCGAGAACCCAGCCAAAGGCTTCCTTCCTGCTG 60
DB 213 GTCCGGCGCCAGCTCCAGTCGAGAACCCAGCCAAAGGCTTCCTTCCTGCTG 272
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCCTGCCCT 120
DB 273 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCCTGCCCT 332
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCCGAG 180
DB 333 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCCGAG 392
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCATTTCCTCCAG 240
DB 393 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCATTTCCTCCAG 452
QY 241 CAACACCCGCGC 251
DB 453 CAACACCCGCGC 463

RESULT 9

AW960290 585 bp mRNA EST 01-JUN-2000
LOCUS EST372361 MAGE resequences, MAGF Homo sapiens cDNA, mRNA sequence.
ACCESSION AW960290
VERSION AW960290.1 GI:8149974
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 585)
 AUTHORS Hegde, P., Qi, R., Aherthy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL: Unpublished (2000)

COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: johnq@tigr.org

Plate: 144

Seq primer: Reversal

FEATURES Location/Qualifiers

1..585

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resesquences, MAGF"

/note="Vector: pBluescriptSkm"

BASE COUNT 134 a 190 c 148 g 113 t

ORIGIN

Query Match 93.8%; Score 242; DB 97; Length 585;

Best Local Similarity 96.1%; Pred. No. 1.2e-51;

Matches 248; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGCTGCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 60

DB 19 GGCGGGCGCCAGCTGCAGTCGAGAACCCAGCCAAAGGCTTGTATTTCTGCTG 78

QY 61 CTACCATCTCTTCTGCCAGATCCTGATGGCTGAGAGAGGTGTGCCGGCGCCCTGCCT 120

DB 79 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAGAGAGGTGTGCCGGCGCCCTGCCT 138

QY 121 CCAGAGACGCCCTTACGCCGATCCCTGCGGCCACCCCTGTGTCGCCCTGCTGAG 180

DB 139 CCAGAGACGCCCTTACGCCGATCCCTGCGGCCACCCCTGTGTCGCCCTGCTGAG 198

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCTCCAG 240

DB 199 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACACTCTGGACCTGAGCACTTCTCTCCAG 258

QY 241 CACACCCCGCGCCTTC 258

DB 259 CAACACCCCGCGCCTTC 276

RESULT 10

BE563592

LOCUS

DEFINITION 60134706f1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3688425 5', mRNA sequence.

ACCESSION BE563592

VERSION BE563592.1 GI:9807312

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 810)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Place: LLC382 row: k column: 02

High quality sequence stop: 692.

FEATURES

Location/Qualifiers

1..810

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3688425"

/clone_lib="NIH_MGC_39"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pOTB7; Site: 1; XhoI; Site: 2; EcoRI; cDNA made by oligo-dT priming"

Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).

BASE COUNT 176 a 284 c 205 g 145 t

ORIGIN

Query Match 91.8%; Score 236.8; DB 109; Length 810;

Best Local Similarity 98.8%; Pred. No. 2.6e-50;

Matches 249; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GTCGGCGCCAGCTGCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 60

DB 207 GTCGGCGCCAGCTGCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 266

QY 61 CTACCATCTCTTCTGCCAGATCCTGATGGCTGAGAGAGGTGTGCCGGCGCCCTGCCT 120

DB 267 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAGAGAGGTGTGCCGGCGCCCTGCCT 326

QY 121 CCAGAGACGCCCTTACGCCGATCCCTTGGGCCACCCCTGTGTCGCCCTGCTGCGA 179

DB 327 CCAGAGACGCCCTTACGCCGATCCCTTGGGCCACCCCTGTGTCGCCCTGCTGCGA 386

QY 180 GCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCCTCCA 239

DB 387 GCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCCTCCA 446

QY 240 GCAACCCCGCGC 251

DB 447 GCAACCCCGCGC 458

RESULT 11

W52268

LOCUS

DEFINITION

447 bp mRNA EST

W52268

VERSION

W52268.1 GI:1349380

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 447)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 3171 Std Error: 0.00

Seq primer: mob.RECA+BT

High quality sequence stop: 401.

FEATURES
source

Location/Qualifiers
1..447
/organism="Homo sapiens"
/db_xref="GDB:1256850"
/db_xref="taxon:9606"
/clone_lib="NCI-CCAP_Kid3"
/clone_lib="Soares_senescent_fibroblasts_NHMF"
/tissue_type="senescent_fibroblast"
/lab_host="DH10B"
/note="Vector: pT73D (Pharmacia) with a modified polylinker V-type; phagemid; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAGTGGGAGCGCGCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 72 a 188 c 109 g 75 t 3 others
ORIGIN

Query Match 86.4%; Score 223; DB 147; Length 447;
Best Local Similarity 99.2%; Pred. No. 7.9e-47;
Matches 245; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GTCGGCGCCAGTGCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
Db GTCCGGCGCCAG-TGCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 260
QY 61 CTCACATCTCTTCTGCCAGATCCCTGATCGCTGAAGAGGCTGTGCGGCGCCCTGCC 120
Db CTCACATCTCTTCTGCCAGATCCCTGATCGCTGAAGAGGCTGTGCGGCGCCCTGCC 320
QY 121 CCAGAGACGCCCTTAACGGCGCATCCCT-GGCGGCCAGCCCTGTGCTGCCCTCTCGA 179
Db CCAGAGACGCCCTTAACGGCGCATCCCTGGGCGCCACCCCTGTGCTGCCCTCTCGA 380
QY 180 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGCACTTCTCTCA 239
Db GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGCACTTCTCTCA 440
QY 240 GCAACAC 246
Db GCAACAC 447

RESULT 12
AA743819/C
LOCUS
DEFINITION
ob01d04.s1 NCI-CCAP_Kid3 Homo sapiens cDNA clone IMAGE:1322407 3'
similar to SW:1EX1_HUMAN P46695 RADIATION-INDUCIBLE IMMEDIATE-EARLY
GENE IEX-1 ; mRNA sequence.

ACCESSION
VERSION
AA743819
KEYWORDS
EST.
SOURCE
human.

ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 513)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE
JOURNAL
COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
unknown library type
Insert Length: 834 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 384.

FEATURES
source

Location/Qualifiers
1..513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CCAP_Kid3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 124 a 138 c 147 g 104 t
ORIGIN

Query Match 84.2%; Score 217.2; DB 11; Length 513;
Best Local Similarity 98.6%; Pred. No. 2.4e-45;
Matches 219; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 37 GCCAAAGGCTTCTTCTTCTGCTCACCATCGTCTTCTGCCAGATCTCTGATGGCTGAA 96
Db GCCAAAGGCTTCTTCTTCTGCTCACCATCGTCTTCTGCCAGATCTCTGATGGCTCAA 454
QY 97 GAGGTGTGCGGCGCCCTGCTCCAGAGAGCGCCCTTAAGCGCCCATCTTGGCGCCC 156
Db GAGGTGTGCGGCGCCCTGCTCCAGAGAGCGCCCTTAAGCGCCCATCTTGGCGCCC 394
QY 157 ACCCTGTGTCGCCCTCTCTCGAGCCCTTTAATCTGACITTCGAGCCCTCGGACTACGCT 216
Db ACCCTGTGTCGCCCTCTCTCGAGCCCTTTAATCTGACITTCGAGCCCTCGGACTACGCT 334
QY 217 CTGGACCTCAGCACTTCTCTCAGCAACACCGCGCGCCCTTC 258
Db CTGGACCTCAGCACTTCTCTCAGCAACACCGCGCGCCCTTC 292

RESULT 13

H80073
LOCUS
DEFINITION
Yv80gl.r1 Soares melanocyte 2N8M Homo sapiens cDNA clone
IMAGE:249092 5' similar to SP:S33363 S33363 GLY96 PROTEIN - mRNA
sequence.

ACCESSION
VERSION
H80073
KEYWORDS
EST.
SOURCE
human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 212)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE

JOURNAL

COMMENT

The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 192

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: M13RP1

High quality sequence stop: 192.

Location/Qualifiers

1..212

/organism="Homo sapiens"

/db_xref="GDB:3866798"

/db_xref="taxon:9606"

/clone="IMAGE:249092"

/sex="Male"

/clone_lib="Soares melanocyte 2NDHM"

/tissue_type="melanocyte"

/lab_host="DH10B (ampicillin resistant)"

/note="Vector: pT7T3D (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGCTTACCAATCTGAAGTGGAGCGCGCGAGTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M. Fálama Bonaldo. RNA from normal foreskin melanocytes

(FS374) was kindly provided by Dr. Anthony P. Albino."

34 a 81 c 49 g 44 t 4 others

BASE COUNT 276 a 324 c 302 g 168 t

ORIGIN

Query Match 79.6%; Score 205.4; DB 142; Length 212;
Best Local Similarity 97.6%; Pred. No. 2.2e-42;
Matches 206; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 40 AAAAGGCTTCTCTTCTGCTGCTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAG 99

Db 1 AAAAGGATNNCTTTCTGCTGCTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAG 60

QY 100 GGTGTGCGCGCGCCCTGCTCCAGAGAGCGCCCTTAAGCCGCATCCCTGGCGCCACACC 159

Db 61 GGTGTGCGCGCGCCCTGCTCCAGAGAGCGCCCTTAAGCCGCATCCCTGGCGCCACACC 120

QY 160 CCTGTGTCCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTG 219

Db 121 CCTGTGTCCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTG 180

QY 220 GACCTCAGCACTTCTCTCCAGCAACACCGG 250

Db 181 GACCTCAGCACTTCTCTCCAGCAACACCGG 211

RESULT 14

BE878459

LOCUS

DEFINITION

BE878459 1070 bp mRNA EST 27-SEP-2000

601487866f1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890248 5',

mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1070)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: DCTD/DRP/Gazdar

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM660 row: c column: 17

High quality sequence stop: 645.

Location/Qualifiers

1..1070

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3890248"

/clone_lib="NIH_MGC_69"

/tissue_type="large cell carcinoma, undifferentiated"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.1 kb. Library constructed by Life

Technologies."

BASE COUNT 276 a 324 c 302 g 168 t

ORIGIN

Query Match 78.7%; Score 203; DB 137; Length 1070;
Best Local Similarity 98.1%; Pred. No. 1e-41; 0; Indels 5; Gaps 5;
Matches 258; Conservative 0; Mismatches 5; Gaps 5;

QY 1 GTCCGCGCCGAG-CTGCCAGTCGAGGACCGAACCCAGCCCAA-AAGGCTTCTCTTTCTGCG 58

Db 63 GTCCGCGCCGAGTCTGCCAGTCGAGGACCGAACCCAGCCCAAAGAGGCTTCTCTTTCTGCG 122

QY 59 TGC-TCACCATCTCTTCTGCCAGATCCTGATGCTGAGAGGGGTGTCGGCGGCCCTG 117

Db 123 TCATCACCATCTCTTCTGCCAGATCCTGATGCTGAGAGGGGTGTCGGCGGCCCTG 182

QY 118 CTCCAGAGGAC-GCCCTTAACCCGCGCATCCCTGGC-GCCACCCCTGTGTCCCCGCTCC 175

Db 183 CTCCAGAGGAGTCCCTTAACCCGCGCATCCCTGGCTGCCACCCCTGTGTCCCCGCTCC 242

QY 176 TCGAGCCCTTTAATCTGACCTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTTC 235

Db 243 TCGAGCCCTTTAATCTGACCTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTTC 302

QY 236 TCCAGCAACACCGCGCGGCTTC 258

Db 303 TCCAGCAACACCGCGCGGCTTC 325

RESULT 15

AA143404

LOCUS

DEFINITION

AA143404 504 bp mRNA EST 08-NOV-1997

z066e09.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone

IMAGE:591880 5' similar to SW:GL96_MOUSE P46694 IMMEDIATE EARLY

PROTEIN GLY96. ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 504)

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krisman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

White, Y., Wyllie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1695 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 392.

FEATURES

source

1..504

/organism="Homo sapiens"

/db_xref="GDB:4623266"

/db_xref="taxon:9606"

/clone="IMAGE:591880"

/clone_lib="Strategene pancreas (#937208)"

/lab_host="SOUR cells (kanamycin resistant)"

/note="Organ: pancreas; Vector: pBluescript SK-; Site.1:

ECORI; Site.2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Pancreatic adenocarcinoma cell line. Average

insert size: 1.0 Kb; Uni-ZAP XR vector; -5' adaptor

sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT
ORIGIN

80 a 202 c 121 g 92 t 9 others

Query Match 76.7%; Score 197.8; DB 2; Length 504;
Best Local Similarity 90.7%; Pred. No. 2e-40;
Matches 233; Conservative 0; Mismatches 19; Indels 5; Gaps 2;

QY 1 GTCCGGCGCAGTGCCTGAGGAAACCGAACCC-AGCCAAAGGCTTCTCTTCTGCT 59
DB 225 GTCCGGCGCAGTGCCTGAGGAAACCGAACCCAGCCAAAGGCTTCTCTTCTGCT 284
QY 60 GCTACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCGGCCCTGCTGCC 119
DB 285 GCTACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCGGCCCTGCTGCC 344
QY 120 TCCAGAGAGCGCCCTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCC"CGA 179
DB 345 TCCAGAGAGCGCCCTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCC"CGA 404
QY 180 GCCCTTTAATCTGACTTCGGAGCGCTCGG-...ACTACGCTCTGGACCTCAGCAGTTTCC 235
DB 405 GCNCTTTAATCTGACTTCGGAGCGCTTCGGAACCTTCTTGGAACTTAAGCAGCTTCC 464
QY 236 TCCAGACACACCGGCC 252
DB 465 TCCAGACACACCGGCC 481

Search completed: January 30, 2001, 18:47:28
Job time: 28459 sec

CC by RNA polymerase II, the RNA polymerase which transcribes messenger
CC RNA. RNA polymerase II transcription proceeds *in vitro* upon addition
CC of several nuclear fractions designated TFIIA, B, D, F, H, I and J
CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
CC contain a TAP and other TAPS. Purification of TFIID and separation of
CC its components reveals 7 proteins ranging in size from 30-250 kD.
CC Serum raised against the TFIID fraction allowed cloning of the corres.
CC genes from lambda-gt10 expression libraries.
XX
SQ Sequence 1120 BP; 307 A; 291 C; 312 G; 210 T; 0 other;

Query Match	13.0%;	Score 33.6;	DB 17;	Length 1120;
Best Local Similarity	51.0%;	Pred. No. 2.2;		
Matches 104;	Conservative	0;	Mismatches 99;	Indels 1;
Gaps				
QY	34	CCAGCAGAAAGCTTCTCTTCTGCTGCTCCACATCGTCTTCGCGGCAGATCCTGATGCT	93	
Db	866	CCACTCCACGCCAGATGATCTCTCTCTCTCTCGCTGGCAGCTTCCTGCT-CCC	808	
QY	94	GAAGAGGTGTGCGCGCGCCCTGCTCCAGAGGACGCCCTTAAGCGGCATCCTCGTGGC	153	
Db	807	GAACCGGATGCGCCAGCATGTGTGCCACGGCGCGGATCGCTGCCACCTCCATCTTG	748	
QY	154	CCACCCCTGTGTCGCCGTCTCGAGCGCTTTAATCTGACTTCGGAGCCCTCGACATAC	213	
Db	747	AGATCTCCCGCGCGGTGCTCTGCTCTTTAACTCTCTGACCAACCGCGCCCAAGAT	688	
QY	214	GTCTGGACCTCAGCACTTTCCCTC	237	
Db	687	CGCGCGAGCTGCCACCGTTTTTC	664	

RESULT 15

T79601/c
ID T79601 standard; cDNA; 1120 BP.

AC T79601;

D'I' 08-OCT-1997 (first entry)

DE TATA-binding protein associated factor, dTAFFI40 cDNA.

...
KW TATA-binding protein associated factor; TAF; nuclear protein;
KW RNA polymerase transcription; TATA-binding protein; TBP;
KW initiation; ds.

OS *Drosophila* sp.

	Key	Location/Qualifiers
CDS		80..916
		/*tag= a

PN US5637686-A.

10-JUN-1997.
PD

28-JAN-1993; 93US-0013412.

PR 28-JAN-1994; 94US-0188582.

28-JAN-1993; 93US-0013412.
30-JUN-1993; 93US-0087119.

XX
03 MAY 1990, 2003-0046/13.

XX
FA (REC) UNIV CALIFORNIA.

COMAI L., DYNIAČEK B., HOEY T., RUPPERT S., TANASE N.,
TILIAN R.: Wang E. Weinzierl ROJ:

WPI: 1997-319113/29

XX
P-PSDB; W25025.

PF1 Nucleic acids encoding human TATA-binding protein associated factor

PT (TAF) peptide(s) - for production of recombinant peptide(s), used
PT for modulating transcription of TAFs
XX
XX Example 1; Column 61-64; 86pp; English.
XX
CC T79601 encodes Drosophila TATA-binding protein associated factor
CC (TAF) polypeptide, dTAFII40 (mol. weight 40kD). TAF peptides derived
CC from dTAFII30 alpha, dTAFII30 beta, dTAFII40, dTAFII60, dTAFII80,
CC dTAFII110, dTAFII150, and dTAFII250, their human equivalents and
CC nucleic acids encoding them, are used to modulate transcription,
CC including transcription initiation. TAFs are nuclear proteins involved
CC in RNA polymerase I, II and III transcription. The peptides act by
CC binding to a different TAF, an activator, or TBP (TATA-binding protein)
CC or competitively inhibiting association of a TAF domain with another
CC compound, typically a protein like TBP or another TAF, an activator,
CC or DNA.
XX
XX Sequence 1120 BP; 307 A; 291 C; 312 G; 210 T; 0 other;
XX

Query Match 13.0%; Score 33.6; DB 18; Length 1120;
Best Local Similarity 51.0%; Pred. No. 2.2;
Matches 104; Conservative 0; Mismatches 99; Indels 1;

[illegible]

Search completed: January 30, 2001, 21:29:32
Job time: 19258 sec

CC surface molecule 4F2. Antisense DNA or RNA, antibodies, antibody
 CC fragments, and peptide and non-peptide antagonists to LAT1 are useful
 CC as anticancer agents. The present sequence encodes human LAT1, which is
 CC specifically claimed in the present invention.

XX Sequence 4539 BP; 765 A; 1450 C; 1297 G; 998 T; 29 other;

Query Match 13.1%; Score 33.8; DB 21; Length 4539;
 Best Local Similarity 53.0%; Pred. No. 2.6;
 Matches 62; Conservative 2; Mismatches 53; Indels 0; Gaps 0;

QY 8 GCCAGCTGCCAGTCGAGACCGAACCCAGCCCAAGAGGCTTCTTCTGCTGCTCACCA 67
 III | III | III | III | III | III | III | III | III | III | III | III |
 Db 556 gccgggtyccggaggaggaagcaagctgtgctggtggtggtggtgctgctcaagg 615
 III | III | III | III | III | III | III | III | III | III | III | III |
 QY 68 TCGTCTTCTGCCAGATCCTGCTGAAGAGGCTGTGCCGCCGCCCTGCCCTCCAG 124
 III | III | III | III | III | III | III | III | III | III | III | III |
 Db 616 ccgtaactctacagcgtgaaggngcnaacncgggtccaggatgctttgcccgcg 672

RESULT 13

Q70728/c
 ID 070728 standard; cDNA; 1120 BP.

XX AC Q70728;

XX DT 23-MAR-1995 (first entry)

XX DE TATA-binding protein-associated factor dTAFLI40 cDNA.

XX KW TATA-binding protein associated factor; dTAFLI40; ss: screening;
 diagnostic; therapeutic; gene transcription regulation.

XX OS Drosophila.

XX FH Key Location/Qualifiers
 CDS 80..913
 FT /*tag= a

XX WT W09417087-A.

XX PD 04-AUG-1994.

XX PF 28-JAN-1994; 94WO-US01114.

XX PR 28-JAN-1993; 93US-0013412.

XX PR 30-JUN-1993; 93US-0087119.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;

XX PI Tjian R, Wang E, Weinzierl ROJ;

XX WPI: 1994-264019/32.

XX DR P-PSDB; R56490.

XX PT TATA-binding protein associated protein factors - and
 corresponding nucleotide sequence and deriv. antibodies, useful
 in screening, diagnostics and therapeutics

XX PS Disclosure; Page 89-92; 180pp; English.

XX CC The TATA-binding protein associated factor dTAFLI40 (including
 specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors
 CC gene expression associated with human pathology.

XX SQ Sequence 1120 BP; 307 A; 292 C; 311 G; 210 T; 0 other;

Query Match

Best Local Similarity 13.0%; Score 33.6; DB 15; Length 1120;
 Matches 104; Conservative 0; Mismatches 99; Indels 1; Gaps 1;

QY 34 CCAGCCAAAAGGCTTCTCTTTCTGCTGCACCATGCTCTTCTGCGACATCCCTGATGGCT 93
 III | III | III | III | III | III | III | III | III | III | III | III |
 Db 866 CCACTCCAAAGCCAGATGATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 808
 QY 94 GAAGAGGNGTCCGGCGCCCTCCAGAGAGGAGGCCCTTAAGCCGCATCCCTGGCG 153
 III | III | III | III | III | III | III | III | III | III | III | III |
 Db 807 GAACCGGATGCCCGACCGATGCTGCCCGCGCGCGCATCCTGTCCACCTCCATCTTG 748
 QY 154 CCCACCCCTGTCTCCCGCTCTCGAATCTTAATCTGACTTCGAGAGCCCTCGGACTAC 213
 III | III | III | III | III | III | III | III | III | III | III | III |
 Db 747 ACATCTCCCGCGCGCGCTGCTCTGCTCTTAACCTCTGACACACGCCGCCCCAGAT 688
 QY 214 GCTCTGGACCTCAGCACTTTCCTC 237
 III | III | III | III | III | III | III | III | III | III | III | III |
 Db 687 CCGCGGAGCTGCCACCGTTTTC 664

RESULT 14

T42214/c
 ID T42214 standard; cDNA; 1120 BP.

XX AC T42214;

XX DT 27-JAN-1997 (first entry)

XX DE Drosophila TATA-binding protein associated factor dTAFLI60 gene.

XX KW Drosophila; TATA-binding protein; TBP associated factor; TFIID;
 RNA polymerase II; transcription; messenger RNA; nuclear fraction;
 KW holoenzyme; lambda-gt11; expression library; ds.

XX OS Drosophila melanogaster.

XX FH Key Location/Qualifiers
 CDS 80..916
 FT /*tag= a
 FT /product= Drosophila TAFI160

XX PN US5534410-A.

XX PD 09-JUL-1996.

XX PF 28-JAN-1993; 93US-0013412.

XX PR 28-JAN-1994; 94US-0188582;

XX PR 28-JAN-1993; 93US-0013412.

XX PR 30-JUN-1993; 93US-0087119.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;

XX PI Tjian R, Wang E, Weinzierl ROJ;

XX WPI: 1996-333245/33.

XX DR P-PSDB; T42214.

XX PT Screen for cpds. that bind human TATA-binding protein associated
 factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.

XX PS Examples; Column 61-64; 86pp; English.

XX CC This is the nucleotide sequence encoding the Drosophila TATA-binding
 CC protein (TBP) associated factor (TAF) designated TAFI160. The protein
 CC is a component of the TFIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The encoded protein
 CC has an estimated mol. wt. of 60 kD by SDS-PAGE.
 CC The invention relates to purified proteins involved in transcription

CC attack, neurological disorders, e.g. neuroblastoma, glioblastoma or
 CC cancers, or to promote apoptosis in cells, or treat refractory
 CC epilepsy. They can also be used for screening test substances for
 CC IBI biological activity. IBI producing cells can be used
 CC therapeutically to produce IBI in a subject.

XX
 SQ Sequence 2953 BP; 620 A; 906 C; 840 G; 587 T; 0 other;

Query Match 13.5%; Score 34.8; DB 19; Length 2953;
 Best Local Similarity 62.8%; Pred. No. 1.3;
 Matches 54; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 98 AGGTGTGTCGGCGCCCTGCTCCAGAGGAGCCCTTAACGCCGCATCCCTGGCGCCCA 157
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2539 AGGATGGTGCAGCGCCTGTGCTTAGGGGACCCCTTTGTGGCCAGATGCTGACCCCA 2480
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 158 CCCCTGTGTCCCGCTCTCGAGCCC 183
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2479 CCCAGCGCCTGCCCTTCCCGAGCCC 2454

RESULT 11
 Q30849/c
 ID Q30849 standard; cDNA; 3690 BP.

XX AC Q30849;

DT 24-MAR-1993 (first entry)

XX Type III procollagen (prior art).

XX Mutation; pro-alpha1(III); primer; PCR; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 103..3690
 FT FT /*tag= a
 FT FT /label= Type_III_procollagen

PN W09219754-A.

XX 12-NOV-1992.

XX 08-MAY-1992; 92WO-US03866.

XX 08-MAY-1991; 91US-0696607.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Kuivaniemi SH, Prockop DJ, Tromp GC;

XX WPI; 1992-398878/48.

XX P-PSDB; R28916.

XX Kit for detecting genetic pre-disposition for vascular aneurysms
 PT - contains primer to amplify portions of Type III procollagen DNA
 PT and detects mutation in standard procollagen DNA

XX Disclosure; Fig 1A-F; 44pp; English.

XX Example 1 describes the determination of the presence of a mutation
 CC in the pro-alpha1(III) gene. Primers used in PCR are given in
 CC Q30834-48.

XX Sequence 3690 BP; 771 A; 1031 C; 1152 G; 736 T; 0 other;

Query Match 13.3%; Score 34.2; DB 13; Length 3690;
 Best Local Similarity 46.2%; Pred. No. 1.9;
 Matches 114; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 1 GTCGCGCGCCAGCTGCCAGTCGAGAACCCAGCAAAAGCTTCCTCTTCCTG 60
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1201 GTCCAGGGGACCACTTTGAACAGGAGACCTCGAGGTCCCACTTACCTTACACCAAG 1142
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 61 CTCACCATGCTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCGGGCCGCCCTGCT 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1141 GGGATCCAGGAATGCCGCGAGTTCCAGAGGAGCAGGAGGGGTGGTTCAGCATCACTGC 1092
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 121 CCAGAGAGCGCCCTTAACGCCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAG 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1081 CTCGAGCACGTCATTAACCCCGAGCAGCTGGAGGCCAGGAAGTCTTGGCGGTGCTGCT 1022
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 181 CCCTTTAATCTGACTTCGAGAGCCCTCGGACTACGCTCTGAGCTCAAGCTTTCTCTCCAG 240
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1021 CACCAGAGCCCGTCTTGGAGCATGGTTCGAGGAGCTGGATTTCGCCCTTGAGACCAT 962
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 241 CAACACC 247
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 961 TTTCACC 955

RESULT 12

A08372

ID A08372 standard; cDNA; 4539 BP.

XX AC A08372;

DT 12-JUL-2000 (first entry)

XX Human L-type amino acid transporter 1 nucleotide sequence SEQ ID NO:1.
 KW L-type amino acid transporter 1; LAT1; 4F2hc; 4F2 heavy chain;
 KW neutral amino acid transporter; tumour; cancer; proliferation;
 KW cell membrane surface 4F2 molecule; anticancer; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 5'UTR 1..65

FT CDS /*tag= a

FT /*tag= 66..1589

FT /*tag= b

FT /*product= "LAT1"

FT 3'UTR 1590..4474

FT /*tag= c

XX WO200014228-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-JP04789.

XX 03-SEP-1998; 98JP-0249993.

XX 02-SEP-1999; 99JP-0248546.

XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Endou H, Kanai Y;

XX WPI; 2000-256979/22.

XX P-PSDB; Y82488.

XX Neutral amino acid transporter protein which conjugates with cell

PT membrane protein 4F2 and operates independently of sodium ions is

PT useful for screening potential cancer proliferation inhibitors -

XX Claim 9; Page 125-132; 189pp; Japanese.

XX The present invention describes a cell surface protein which is an

CC L-type amino acid transporter-1 (LAT1), which mediates the transport

CC of neutral amino acids, leucine, isoleucine, phenylalanine, methionine,

CC tyrosine, tryptophan, valine and histidine, into the cell independently

CC of sodium ions. The LAT1 protein conjugates with the cell membrane

```

RESULT 9
A02484/c
ID A02484 standard; cDNA; 1000 BP.
XX
AC A02484;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2475.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN W09958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 98WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
WPI: 2000-126369/11.
XX
Polynucleotide library used to determine cancerous states of mammalian
cells -
XX
Claim 1; Page 994; 1097pp; English.
XX
A00010 to A02716 represent polynucleotides isolated from cDNA libraries
constructed from human colon cancer cell lines. The present invention
also describes a method of detecting differentially expressed genes
correlated with a cancerous state of a mammalian cell, comprising
detecting at least one differentially expressed gene product in a test
sample derived from a cell suspected of being cancerous, where detection
of the differentially expressed gene product is correlated with a
cancerous state of the cell from which the test sample was derived.
The polynucleotides sequences can be used in a method for detecting
differentially expressed genes correlated with a cancerous state of a
mammalian cell. The polynucleotides can also be used as probes for
detecting and mapping related genes. They can be used in diagnosis and
prognosis of diseases and disorders (e.g. identification of
pre-metastatic or metastatic cancerous states, stages of cancer, or
responsiveness of cancer to therapy). This is particularly for breast
cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
negative breast cancer, lung cancer, and colon cancer.
XX
Sequence 1000 BP; 21 A; 53 C; 574 G; 22 T; 330 other;
XX
Query Match 14.0%; Score 36.2; DB 21; Length 1000;
Best Local Similarity 44.7%; Pred. NO. 0.42;
Matches 68; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 106 CCGGCGCCCTGCTCCAGAGAGCGCCCTAAGCGCGCATCCCTGGCGCCACCCCTGTG 165
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 803 CCTNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 744

```

```

QY 166 TCCCGCGTCTCGAGCCCTTTAACTCTGACTTCGAGACCCCTCGGACTTACGCTCTGGACCCTC 225
IIIII III IIII IIII IIII IIII IIII IIII IIII
DB 743 CCCCCCCCCCTCCCCCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNC 684
QY 226 AGCACTTTCTCCAGCAACACCCGCGCCCTT 257
II III IIII IIII IIII IIII IIII IIII
DB 683 CCCCCCNCNCCTGCCNCCNCNCNCNCNCNCNCCT 652
RESULT 10
V62462/c
ID V62462 standard; cDNA; 2953 BP.
XX
AC V62462;
XX
DT 15-FEB-1999 (first entry)
XX
DE Rat transcription factor islet-brain 1 (IB1) cDNA.
XX
KW IB1; islet-brain 1; transcription factor; rat; diabetes;
KW dementia; Parkinson's disease; Alzheimer's disease; epilepsy;
KW neuroblastoma; glioblastoma; apoptosis; cancer; autoimmune disease;
KW systemic lupus erythematosus; myocardial infarction; ischaemia;
KW diagnosis; therapy; ss.
XX
OS Rattus sp.
XX
Key Location/Qualifiers
FT CDS 108..2252
FT /tag= a
XX
PN W09844106-A1.
XX
PD 08-OCT-1998.
XX
PF 02-APR-1998; 98WO-GR00972.
XX
PR 15-MAY-1997; 97GB-0009920.
PR 03-APR-1997; 97GB-0006731.
XX
(KIDD/) KIDDLE S J.
PA (NICO/) NICOD P.
PA (WAEB/) WAEBER G.
XX
Bonny C, Waeber G;
XX
WPI: 1998-568278/48.
P-PSDB; W80601.
XX
New isolated transcription factor islet-brain 1 - used to develop
products for treating e.g. diabetes, neurodegenerative disorders,
cancers, autoimmune disease, heart disease or epilepsy
XX
Claim 13; Fig 1A; 111pp; English.
XX
This cDNA sequence includes an open reading frame encoding rat
islet-brain 1 (IB1, see W80601), a transcriptional activator that
is involved in the control of the GLUT2 and insulin genes by
interacting with homologous cis-regulatory elements of the GLUT2
and insulin gene promoters. IB1 cDNA was isolated from a rat
differentiated insulinoma INS-1 cDNA using GT11 oligonucleotides
(see V62464) as probe; GT11 is a cis element located in the proximal
region of the GLUT2 promoter. IB1 is abundantly expressed in the
pancreatic islets in the brain. Human IB1 polynucleotide (see
V62463) and polypeptide (see W80602) are also claimed. IB1
polypeptides, nucleic acids, agonists and antagonists can be used in
the treatment or diagnosis of diabetes, neurodegenerative disorders
or cancers, to inhibit cells from undergoing apoptosis, to treat
dementia, Parkinsonism, Alzheimer's disease, neuronal disabilities
such as speech disorders and memory alteration, autoimmune diseases
affecting the CNS such as systemic lupus erythematosus, diabetes,
heart diseases such as myocardial infarct and ischemia or brain

```


PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Belt JA, Crawford CR, Patel DH;

XX WPI: 1998-594576/50.

XX New equilibrative nucleoside transport protein insensitive to
PT nitrobenzylthio-purine riboside - useful for, e.g. identifying
PT specific therapeutic nucleoside analogues and in gene therapy to
PT protect transduced cells against ablative chemotherapy

XX Example 3; Pages 76-79; 114pp; English.

XX This represents the 5' UTR (untranslated region) of the DNA that
CC encodes an equilibrative nucleoside transport protein (iENTP), which
CC is insensitive to nitrobenzylmercaptopyrine riboside (NBMPR). Cells
CC transformed with a construct containing the iENTP nucleic acid can be
CC used to produce the protein recombinantly. iENTP is used to identify
CC specific ligands (particularly antiviral and antitumour nucleoside
CC analogues that are preferentially transported into cells) and to raise
CC antibodies. Cells in which iENTP provides all available transport
CC activity are used: (a) to identify permeants of iENTP and (b) to screen
CC specific inhibitors of iENTP (potential drugs). Fragments of the iENTP
CC nucleic acid are used, as probes, primers, antisense molecules, and
CC ribozymes for therapy or diagnosis, and knockout mice in which both
CC alleles encoding iENTP contain an inactivating defect are also useful for
CC drug screening. Cells that have been transduced with iENTP nucleic acid
CC ex vivo are used particularly for cancer chemotherapy. Vectors in which
CC the iENTP gene is linked to a heterologous gene (e.g. encoding adenosine
CC deaminase and factor VIII) can be used for gene therapy.

XX Sequence 6354 BP; 1322 A; 1797 C; 1873 G; 1355 T; 7 other;

Query Match 15.0%; Score 38.6; DB 19; Length 6354;
Best Local Similarity 65.9%; Pred. NO. 0.14; Indels 0; Gaps 0;
Matches 56; Conservative 0; Mismatches 29;

QY 100 GGTGTGCGCGCCCTGCTCCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACC 159

DB 2461 ggtgagagcgctgcctgcctgcctctgcgagcgagcttcattgagcgctcc 2520

QY 160 CTGTGTCCCGCCCTCTCCAGCGCCT 184.

DB 2521 cctgcgccccctgcctccagcgccct 2545

RESULT 6

ID Z51683

XX AC

XX Z51683;

XX 04-JUL-2000 (first entry)

XX Human cyclic nucleotide-associate protein-2 (CNAP-2) cDNA.

XX Cyclic nucleotide-associate protein-2; CNAP-2; human; cytosolic;
KW anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;
KW immunomodulatory; anti-asthmatic; anti-anaemic; anti-diabetic; diagnosis;
KW anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;
KW anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;
KW anti-infertility; anti-allergic; vasotropic; immunosuppressive;
KW hypotensive; gene therapy; prevention; treatment; arteriosclerosis;
KW cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;
KW neurological; vision; reproductive; smooth muscle; ss.

OS Homo sapiens.

XX Key

XX Location/Qualifiers

XX 31..4014

XX /tag= a

XX /product= "Human CNAP-2 protein"

FT /note= "Shares 24% identity to Aquifex pyrophilus
FT esterase 28LC"
FT 31..132
FT /tag= b
FT 133..4011
FT /tag= c
FT /product= "Mature CNAP-2 protein"
FT 136..165
FT /tag= d
FT /bound_moiety= "Primer or Probe"
FT /note= "Useful for amplification or hybridisation
FT techniques"

XX W0200014248-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US20287.

XX 04-SEP-1998; 98US-0148904.

XX (INCY-) INCYTE PHARM INC.

XX Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;

XX WPI: 2000-256994/22.

XX P-PSDB; Y70474.

XX Isolated cyclic nucleotide associated proteins useful for preventing,
PT diagnosing and treating cell proliferative, autoimmune/inflammatory,
PT neurological, vision, reproductive and smooth muscle disorders -

XX Claim 9; Page 70-71; 78pp; English.

XX The present sequence is the cDNA encoding human cyclic nucleotide
CC associated protein-2 (CNAP-2), identified in Incyte clone 3149674,
CC that is isolated from ADREN004 cDNA library. It is expressed in
CC nervous, reproductive, cardiovascular and haematopoietic/immune tissues.
CC CNAP sequences may be used for prevention, treatment and diagnosis of
CC diseases associated with altered CNAP expression such as, cell
CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia,
CC lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/
CC inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple
CC sclerosis and psoriasis), neurological disorders (e.g. epilepsy,
CC Alzheimer's/Parkinson's disease and strokes), vision disorders (e.g.
CC conjunctivitis, glaucoma, cataracts and retinitis pigmentosa),
CC reproductive disorders (e.g. infertility, uterine fibroids, ectopic
CC pregnancies and impotence) and smooth muscle disorders (e.g. angina,
CC anaphylactic shock, Kearns-Sayre syndrome and hypertension). It can also
CC be used for gene therapy.

XX Sequence 4228 BP; 793 A; 1365 C; 1324 G; 746 T; 0 other;

Query Match 14.8%; Score 38.2; DB 21; Length 4228;
Best Local Similarity 54.7%; Pred. NO. 0.17;
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 75 CTGCAGATCCTGATGGCTGAGAGGGGTGCGCGCGCCCTGCTCCAGAGAGCGCCCC 134

DB 1185 ctccgacttcgacatggcctatgagcggtggtccgtccctgcaggaagagcc 1244

QY 135 TAACGCCGATCCCTGGCGCCGCCACCCCTGTGTCCCGCGCTTCGAGCGCCCTTTAATCTGAC 194

DB 1245 ctccgggggggtccctgcgagccccccgcctcgcgacccccactcaggaagcctcgagcagcc 1304

QY 195 TTCGAGCGCCCTCGGACTAC 213

DB 1305 ggcaggcgctgtgaatc 1323

RESULT 7

V69706

DT 15-FEB-1999 (first entry)
XX EST clone BY66.
XX
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
XX gene therapy; ss.
XX
OS Homo sapiens.
XX
XX WO9845436-A2.
PN
XX 15-OCT-1998.
PD
XX
XX 10-APR-1998; 98WO-US06955.
PF
XX
XX 10-APR-1997; 97US-0838821.
PR
XX
XX (GEM) GENETICS INST INC.
PA
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX WPI; 1999-070077/06.
DR
XX New polynucleotides encoding human secreted proteins - derived from
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
XX Claim 1; Page 164; 618pp; English.
PS
XX The present sequence represents a human expressed sequence tag (EST).
XX The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
XX Sequence 193 BP; 35 A; 39 C; 68 G; 51 T; 0 other;
SQ

Query Match 32.6%; Score 84; DB 20; Length 193;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTTC 234
DB 193 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTTC 134
QY 235 CTCGAGCAACACCCGCCCTTC 258
DB 133 CTCGAGCAACACCCGCCCTTC 110
RESULT 4
ID Q77534 standard; DNA; 297 BP.
XX
XX Q77534;
AC
XX
XX 23-SEP-1994 (first entry)
DT
XX Human genome fragment.
DE
XX Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.

Query Match 32.6%; Score 84; DB 20; Length 193;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTTC 234
DB 193 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTTC 134
QY 235 CTCGAGCAACACCCGCCCTTC 258
DB 133 CTCGAGCAACACCCGCCCTTC 110
RESULT 4
ID Q77534 standard; DNA; 297 BP.
XX
XX Q77534;
AC
XX
XX 23-SEP-1994 (first entry)
DT
XX Human genome fragment.
DE
XX Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.

XX Homo sapiens.
OS
XX WO9401548-A.
PN
XX 20-JAN-1994.
PD
XX
XX 13-JUL-1993; 93WO-GB01467.
PF
XX
XX 13-JUL-1992; 92GB-0014857.
PR
XX (MEDI-) MEDICAL RES COUNCIL.
PA
XX Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
XX WPI; 1994-035056/04.
DR
XX New nucleic acid fragment encoding gene products - can be used
XX for genetic analysis and mapping
PT
XX Claim 1; Page 575-576; 616pp; English.
PS
XX Human nucleic acid fragments, isolated from brain, adrenal tissue,
XX the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (Q76401-Q77613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
XX
XX Sequence 297 BP; 66 A; 117 C; 61 G; 52 T; 1 other;
SQ

Query Match 17.1%; Score 44; DB 15; Length 297;
Best Local Similarity 77.9%; Pred. No. 0.0026;
Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 2 TCGGCGCCGAGTCCGAGCAACCGACCGCAAGGCTTCTTCTCTCTGCTGC 61
DB 225 tcggcgccagtgccagtttaaggagacaaacctgccagaggctcttctcggattga 284
QY 62 TCACCATC 69
DB 285 tcggccatc 292
RESULT 5
ID V69699 standard; DNA; 6354 BP.
XX
XX V69699;
AC
XX 08-FEB-1999 (first entry)
DT
XX 5' UTR sequence of iENTP DNA.
DE
XX
XX Equilibrative nucleoside transport protein; iENTP; NBMPR; transport;
KW nitrobenzylmercaptopyrine riboside; antiviral; antitumour; screening;
KW inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;
KW adenosine deaminase; factor VIII; UTR; untranslated region; ds.
XX
XX Homo sapiens.
OS
XX WO9846749-A1.
PN
XX 22-OCT-1998.
PD
XX
XX 10-APR-1998; 98WO-US07283.
PF
XX
XX 09-APR-1998; 98US-0058389.
PR
XX 11-APR-1997; 97US-0838845.
XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 21:29:29 ; Search time 77.57 Seconds
(without alignments)
1249.465 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468

Perfect score: 258
Sequence: 1 GTCCGGCCCGCCAGTCCAGT.....AGCAACACCCGCCGCCATC 258.

Scoring table: IDENTITY_NUC %
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36:*

1:	/cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	258	100.0	1228 18 T94471	Human Fchd605 gene
2	258	100.0	1228 21 Z50711	Nucleotide sequenc
3	84	32.6	193 20 V89271	EST clone BY66. H
4	44	17.1	297 15 Q77534	Human genome fragm
5	38.6	15.0	6354 19 V69699	5' UTR sequence of
6	38.2	14.8	4228 21 Z51683	Human cyclic nucle
7	37.6	14.6	170 19 V69706	Nucleotide sequenc
8	36.4	14.1	1526 20 Z42235	Human normal bladd
9	36.2	14.0	1000 21 A02484	Human colon cancer
10	34.8	13.5	2953 19 V62462	Rat transcription
11	34.2	13.3	3690 13 Q30849	Type III procollag
12	33.8	13.1	4539 21 A08372	Human L-type amino

C 13	33.6	13.0	1120	15	Q70728
C 14	33.6	13.0	1120	17	T42214
C 15	33.6	13.0	1120	18	T79601
C 16	33.6	13.0	3765	13	Q31890
C 17	33.6	13.0	5460	17	T16508
C 18	33	12.8	1072	13	Q24285
C 19	33	12.8	11722	19	V34455
C 20	33	12.8	114955	20	X53491
C 21	32.8	12.7	513	20	X91237
C 22	32.8	12.7	514	20	X91400
C 23	32.8	12.7	1450	21	Z47131
C 24	32.8	12.7	8169	19	V26609
C 25	32.4	12.6	201	8	N70194
C 26	32.4	12.6	201	8	N70195
C 27	32.4	12.6	2681	20	X58429
C 28	32.2	12.5	1524	21	A08393
C 29	32	12.4	742	21	Z45675
C 30	32	12.4	1459	21	A02528
C 31	32	12.4	3410	19	V61201
C 32	32	12.4	3410	19	V58586
C 33	32	12.4	3410	21	A06349
C 34	32	12.4	34094	20	Z30163
C 35	31.8	12.3	599	19	V43880
C 36	31.8	12.3	1321	21	A37128
C 37	31.6	12.2	1586	19	V34287
C 38	31.6	12.2	1907	19	V34314
C 39	31.6	12.2	2074	19	V32932
C 40	31.6	12.2	2074	19	V32933
C 41	31.4	12.2	595	14	Q50041
C 42	31.4	12.2	1335	13	Q23295
C 43	31.4	12.2	13842	21	Z87297
C 44	31.4	12.2	36778	21	Z87318
C 45	31.4	12.2	37948	21	Z87285

ALIGNMENTS

RESULT	1
ID	T94471 standard; cDNA; 1228 BP.
AC	T94471;
XX	
DT	03-MAR-1998 (first entry)
XX	
DE	Human Fchd605 gene differentially regulated in monocytes.
XX	
KW	Fchd602 gene; differential expression; monocyte; human;
KW	foam cell; cardiovascular disease; atherosclerosis; pschaemia;
KW	reperfusion; hypertension; restenosis; arterial inflammation;
KW	therapy; diagnosis; drug screening; marker; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key
ET	CDS
ET	Location/Qualifiers
ET	1..471
XX	/tag= a
PN	W09730065-A1.
XX	
PD	21-AUG-1997.
XX	
PF	14-FEB-1997; 97WO-US02291.
XX	
PR	13-FEB-1997; 97US-0799910.
PR	16-FEB-1996; 96US-0011787.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Falb DA;
XX	
DR	WPI; 1997-424966/39.

TATA-binding prote
Drosophila TATA-bi
TATA-binding prote
FMR-1 gene for det
Vector pAC3A1 cont
Epo:IL-3 Flex, rec
Human MHC class I
Human adenosine A1
T. gondii immunoge
Pseudomonas aerugi
Actinomadura hibis
Signal portion of
Streptomyces prote
Thermophilus therm
Human L-type amino
cDNA sequence of a
Human colon cancer
Full length cDNA s
Prostate tumour sp
Human immunogenic
Complete nucleotid
Mycobacterial meth
Human PRO1781 (UNQ
Human secreted pro
Human cyclin-depen
Human cyclin-depen
ICP34.5 fragment.
HSV-1 (F) ICP34.5
S. venezuelae macr
S. venezuelae pik
S. venezuelae pik

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 10:53:09 ; Search time 1450.12 Seconds
(without alignments)
5934.126 Million cell updates/sec

Title: US-08-799-910-9
Perfect score: 1228
Sequence: 1 ATGTGTCACCTCTCGCAGCTG.....ANAAAAAATTCGAG 1228

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:

1: gb_estl1:
2: gb_est2:
3: gb_est3:
4: gb_est4:
5: gb_est5:
6: gb_est6:
7: gb_est7:
8: gb_est8:
9: gb_est9:
10: gb_est10:
11: gb_est11:
12: gb_est12:
13: gb_est13:
14: gb_est14:
15: gb_est15:
16: gb_est16:
17: gb_est17:
18: gb_est18:
19: gb_est19:
20: gb_est20:
21: gb_est21:
22: gb_est22:
23: gb_est23:
24: gb_est24:
25: gb_est25:
26: gb_est26:
27: gb_est27:
28: gb_est28:
29: gb_est29:
30: gb_est30:
31: gb_est31:
32: gb_est32:
33: gb_est33:
34: gb_est34:
35: gb_est35:
36: gb_est36:
37: gb_est37:
38: gb_est38:
39: gb_est39:
40: gb_est40:
41: em_estba:
42: em_estfun:
43: em_esthum1:

44: em_esthum2:
45: em_esthum3:
46: em_esthum4:
47: em_esthum5:
48: em_esthum6:
49: em_esthum7:
50: em_esthum8:
51: em_esthum9:
52: em_esthum10:
53: em_esthum11:
54: em_esthum12:
55: em_esthum13:
56: em_esthum14:
57: em_esthum15:
58: em_esthum16:
59: em_esthum17:
60: em_esthum18:
61: em_esthum19:
62: em_esthum20:
63: em_estin1:
64: em_estin2:
65: em_estin3:
66: em_estin4:
67: em_estov1:
68: em_estov2:
69: em_estpl1:
70: em_estpl2:
71: em_estpl3:
72: em_estpl4:
73: em_estpl5:
74: em_estro1:
75: em_estro2:
76: em_estro3:
77: em_estro4:
78: em_estro5:
79: em_estro6:
80: em_estro7:
81: em_estro8:
82: em_estro9:
83: em_estro10:
84: em_estro11:
85: em_estro12:
86: em_estro13:
87: gb_est41:
88: gb_est42:
89: gb_est43:
90: gb_est44:
91: gb_est45:
92: gb_est46:
93: gb_est47:
94: gb_est48:
95: gb_est49:
96: gb_est50:
97: gb_est51:
98: gb_est52:
99: gb_est53:
100: gb_est54:
101: gb_est55:
102: gb_est56:
103: gb_est57:
104: gb_est67:
105: gb_est68:
106: gb_est69:
107: gb_est70:
108: gb_est71:
109: gb_est72:
110: gb_est73:
111: gb_est74:
112: em_esthum21:
113: em_esthum22:
114: em_esthum23:
115: em_estom1:
116: em_estom2:

117: em_estpl6:*
 118: em_estpl7:*
 119: em_estpl8:*
 120: em_estpl9:*
 121: em_estpl10:*
 122: em_estpl11:*
 123: em_estpl12:*
 124: em_estpl13:*
 125: em_estpl14:*
 126: em_estpl15:*
 127: em_estpl16:*
 128: em_estpl17:*
 129: em_estpl18:*
 130: em_estpl19:*
 131: em_estpl20:*
 132: em_estpl21:*
 133: em_estpl22:*
 134: em_estpl23:*
 135: em_estpl24:*
 136: em_estpl25:*
 137: em_estpl26:*
 138: em_estpl27:*
 139: em_estpl28:*
 140: em_estpl29:*
 141: em_estpl30:*
 142: em_estpl31:*
 143: em_estpl32:*
 144: em_estpl33:*
 145: em_estpl34:*
 146: em_estpl35:*
 147: em_estpl36:*
 148: em_estpl37:*
 149: em_estpl38:*
 150: em_estpl39:*
 151: em_estpl40:*
 152: em_estpl41:*
 153: em_estpl42:*
 154: em_estpl43:*
 155: em_estpl44:*
 156: em_estpl45:*
 157: em_estpl46:*
 158: em_estpl47:*
 159: em_estpl48:*
 160: em_estpl49:*
 161: em_estpl50:*
 162: em_estpl51:*
 163: em_estpl52:*
 164: em_estpl53:*
 165: em_estpl54:*
 166: em_estpl55:*
 167: em_estpl56:*
 168: em_estpl57:*
 169: em_estpl58:*
 170: em_estpl59:*
 171: em_estpl60:*
 172: em_estpl61:*
 173: em_estpl62:*
 174: em_estpl63:*
 175: em_estpl64:*
 176: em_estpl65:*
 177: em_estpl66:*
 178: em_estpl67:*
 179: em_estpl68:*
 180: em_estpl69:*
 181: em_estpl70:*
 182: em_estpl71:*
 183: em_estpl72:*
 184: em_estpl73:*
 185: em_estpl74:*
 186: em_estpl75:*
 187: em_estpl76:*
 188: em_estpl77:*
 189: em_estpl78:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	748.2	60.9	951	107	BE383865	BE383865 601273192
2	621.4	50.6	765	106	BE275966	BE275966 601120727
3	604.2	49.2	762	110	BE621346	BE621346 601493736
4	590.4	48.0	916	137	BE882130	BE882130 601505227
5	589.4	48.0	603	106	BE273386	BE273386 601143493
6	588.4	47.9	810	109	BE563592	BE563592 601334706
7	585.4	47.6	585	137	BE910077	BE910077 601502143
8	581.2	47.3	586	89	AW340002	AW340002 hc91d03.x
9	579.6	47.2	598	93	AW662348	AW662348 hi25c07.x
10	567.2	46.2	614	110	BE621600	BE621600 601493736
11	561.6	45.7	588	97	AW956285	AW956285 EST368355
12	557.4	45.4	617	38	AV690660	AV690660 AV690660
13	544.8	44.4	548	88	AW263252	AW263252 xn79g02.x
14	540.8	44.0	1070	137	BE878459	BE878459 601487866
15	519.4	42.3	521	25	AI800828	AI800828 wgl1f12.x
16	517.4	42.1	544	107	BE384949	BE384949 601276496
17	515.6	42.0	523	18	AI302813	AI302813 qn58g09.x
18	514.8	41.9	585	97	AW960290	AW960290 EST372361
19	513.4	41.8	518	12	AA806234	AA806234 oe29h02.s
20	512.4	41.7	513	40	AW150706	AW150706 xg38a10.x
21	506.2	41.2	524	22	AI566501	AI566501 tr60f07.x
22	505.4	41.0	509	9	AA582942	AA582942 nn80c01.s
23	503.4	41.0	521	23	AI685453	AI685453 tt88d12.x
24	501.6	40.8	510	89	AW337920	AW337920 xw63a07.x
25	496.4	40.4	553	142	N32077	N32077 yw97d08.s1
26	494.8	40.3	1155	135	BE786915	BE786915 601477783
27	487.6	39.7	548	96	AW887044	AW887044 RC1-OT008
28	486.8	39.6	535	25	AI802925	AI802925 tj58a07.x
29	486.6	39.6	494	15	AI022951	AI022951 ow66g09.s
30	482.4	39.3	484	19	AI336188	AI336188 qt44c12.x
31	481.4	39.2	491	39	AW081284	AW081284 xc30c01.x
32	481.4	39.2	492	10	AA643850	AA643850 np26f08.s
33	478.4	39.0	480	96	AW874420	AW874420 hq04a02.x
34	472.4	38.5	496	16	AI160053	AI160053 qc08a02.x
35	472.4	38.4	475	8	AA512938	AA512938 nh91b09.s
36	470.8	38.3	487	23	AI687081	AI687081 tp92h08.x
37	467.8	38.0	701	135	BE735733	BE735733 601304556
38	464.6	37.8	483	16	AI091958	AI091958 qa59d10.s
39	461.4	37.6	466	39	AW051587	AW051587 wv87h05.x
40	458.4	37.3	460	17	AI185199	AI185199 qe35h09.s
41	458.4	37.3	460	24	AI738521	AI738521 w133a06.x
42	458.4	37.3	465	13	AA84985	AA84985 am35e07.s
43	453.6	36.9	461	12	AA812286	AA812286 nr82f07.s
44	451.6	36.8	459	147	W60982	W60982 zc98h09.s1
45	450.8	36.6	469	24	AI761431	AI761431 wq65e08.x

ALIGNMENTS

RESULT 1

LOCUS BE383865

DEFINITION 601273192F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3614255 5', mRNA sequence.

ACCSSION BE383865

VERSION BE383865.1

KEYWORDS EST, human.

BE383865

601273192F1 NIH_MGC_20

Homo sapiens cDNA clone

IMAGE:3614255 5',

951 bp

EST

21-JUL-2000

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM276 row: g column: 24
High quality sequence stop: 747.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

1..951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3614255"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
210 a 324 c 253 g 164 t

BASE COUNT

ORIGIN

Query Match 60.9%; Score 748.2; DB 107; Length 951;
Best Local Similarity 91.5%; Pred. No. 4.8e-180;
Matches 848; Conservative 0; Mismatches 73; Indels 6; Gaps 5;
QY 2 TGTGTCACTTCGACGTCGCCACCCGACCATGACCATCTCTGCAGGCGCCGACCCCGGCC 61
DB 1 TGTGTCACTTCGACGTCGCCACCCGACCATGACCATCTCTGCAGGCGCCGACCCCGGCC 60
QY 62 CTTCCACCATCCCGGACCCCGGGGGTCCGGTCTCTGAGATCTTCACCTTCGACCCCTC 121
DB 61 CTTCCACCATCCCGGACCCCGGGGGTCCGGTCTCTGAGATCTTCACCTTCGACCCCTC 120
QY 122 TCCCGGAGCCCGCAGCGGCCCTCGCGGGCGCCGAGCGCTCTCGCGGCGCACCGAAAGC 181
DB 121 TCCCGGAGCCCGCAGCGGCCCTCGCGGGCGCCGAGCGCTCTCGCGGCGCACCGAAAGC 180
QY 182 GCAGCGCGAGGGTTCTTACCTCTGAGTGTCTCGCGGCCGACGTCGCCAGTCGAGAAACCGA 241
DB 181 GCAGCGCGAGGGTTCTTACCTCTGAGTGTCTCGCGGCCGACGTCGCCAGTCGAGAAACCGA 240
QY 242 ACCGAGCCAAAGGCTCTCTTCTGCTGTCTCACCATCTGCTTCTGCGCAGATCCTGATGG 301
DB 241 ACCGAGCCAAAGGCTCTCTTCTGCTGTCTCACCATCTGCTTCTGCGCAGATCCTGATGG 300
QY 302 CTGAAGAGGCTGTGCGGGCGCTCTGCTGCTCAGAGGACGCCCTTAACGCGGCATCCCTGG 361
DB 301 CTGAAGAGGCTGTGCGGGCGCTCTGCTGCTCAGAGGACGCCCTTAACGCGGCATCCCTGG 360
QY 362 CGGCCACCCCTGTGTCCCGCTCTCTCGAGGCCCTTTAATCTCACTTCGAGGCCCTCGGACT 421
DB 361 CGGCCACCCCTGTGTCCCGCTCTCTCGAGGCCCTTTAATCTCACTTCGAGGCCCTCGGACT 420
QY 422 ACGCTCTGGACCTCAGGACTTTCTCTCAGCAACACCCGGCGGCTTCTTAAGTGTGACTCC 481
DB 421 ACGCTCTGGACCTCAGGACTTTCTCTCAGCAACACCCGGCGGCTTCTTAAGTGTGACTCC 480

QY 482 CCGACTCCCAAAAGAAATCCGAAAACACCAAAACACACAGCGGTACCTGGTGGCC 541
DB 481 CCGACTCCCAAAAGAAATCCGAAAACACCAAAACACACAGCGGTACCTGGTGGCC 540
QY 542 GAGAGCGTATCCCAACTCGGACTTCCGAGGCAACTTGAATCAGAACACTACAGCGGAG 601
DB 541 GAGAGCGTATCCCAACTCGGACTTCCGAGGCAACTTGAATCAGAACACTACAGCGGAG 599
QY 602 AGCCACCCCGTGTCTTGAGCGGGGACCGAGGCGCAGAGAGCGGCGCATAGACCG 661
DB 600 AGCCACCCCGTGTCTTGAGCGGGGACCGAGGCGCAGAGAGCGGCGCATAGACCG 659
QY 662 AGGCACAGCCAGC-TGGGGCTAGGCGCGTGGGAAGGAGAGCGCTCTTAAATTTATTTCT 720
DB 660 AGGCACAGCCAGCTTGGGGCTAGGCGCGTGTGAAGGAGAGCGGTATATATCT 719
QY 721 TATTCCTCTCTAAATTAATTTATATG--TATTATCTAGCTCTCTCTAGGTGATGAGAT 778
DB 720 TATAGCTCTCTAAATTAATTTATATG--TATTATCTAGCTCTCTCTAGGTGATGAGAT 779
QY 779 GTGAGCTAATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTGTAAT 838
DB 780 GTGCGGTACAAATTA-TCAAACCTTATGCAA--GGTGTGACATGTTCCCTCTGTGAAAGC 837
QY 839 GCAGGTCTCTTGTATTTATTTAGCTTTGTGGGACTTGTGGAAGCAGGACACCTTGGAACT 898
DB 838 AGGCACAGTAATTAATGAGCCTGGGACTTGTGGAGCGCAAAACCGGAACTTGGAAAGT 897
QY 899 GCGGCAAGTAGGAGGAAGAAATGGGA 925
DB 898 AGGAGCAACCGCGCGCAGATGGCGA 924
RESULT 2
BE275966 765 bp mRNA 13-JUL-2000
LOCUS 601120727F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2967534 5',
DEFINITION mRNA sequence.
ACCESSION BE275966
VERSION BE275966.1 GI:9150928
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 765)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM276 row: e column: 07
High quality sequence stop: 7
High quality sequence start: 699.
Location/Qualifiers
1..765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2967534"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'

cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 112 a 241 c 154 g 96 t

ORIGIN

Query Match 48.08; Score 589.4; DB 106; Length 603;
Best Local Similarity 99.1%; Pred. No. 1.3e-139;
Matches 601; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 35 CCATCTCTGACGGCCGACCGCCCTCCACCATCCCGGACCCCGGGGCTCCG 94
Db 1 CCATCTCTGACGGCCGACCGCCCTCCACCATCCCGGACCCCGGGGCTCCG 60

QY 95 GTCTTGAGATCTTACCTTCGACCTCTCCCGAGCCCGGAGCCCGCCGCGGCC 154
Db 61 GTCTTGAGATCTTACCTTCGACCTCTCCCGAGCCCGGAGCCCGCCGCGGCC 120

QY 155 CCAGCGCTCTCGGGGACCGAAGCGGACCGGAGGTTCTTACCTTCGAGTGGTCC 214
Db 121 CCAGCGCTCTCGGGGACCGAAGCGGACCGGAGGTTCTTACCTTCGAGTGGTCC 180

QY 215 GGGCCGAGTTCGAGGACCGAAGCGGACCGGAGGTTCTTACCTTCGAGTGGTCC 274
Db 181 GGGCCGAGTTCGAGGACCGAAGCGGACCGGAGGTTCTTACCTTCGAGTGGTCC 240

QY 275 CCATCTCTTCTGCCAGATCTTGATGGCTGAAGAGGCTTTCGCGGCCCTGCCCTCCAG 334
Db 241 CCATCTCTTCTGCCAGATCTTGATGGCTGAAGAGGCTTTCGCGGCCCTGCCCTCCAG 300

QY 335 AGGACGCCCTAACGCGGATCCCTGGCGCCACCCCTGTGTCC--CCGCTCTCGAGCCC 393
Db 301 AGGACGCCCTAACGCGGATCCCTGGCGCCACCCCTGTGTCTTCGCGGCCCTGCCCTCCAG 360

QY 394 TTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACTCTGAGCTTCTTCTCCAGCAA 453
Db 361 TTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACTCTGAGCTTCTTCTCCAGCAA 420

QY 454 CACCGCGCCCTTCTAAGTCTGACTCTGAGCTCTGAGCTCTGAGCTTCTTCTCCAGCAA 513
Db 421 CACCGCGCCCTTCTAAGTCTGACTCTGAGCTCTGAGCTCTGAGCTTCTTCTCCAGCAA 480

QY 514 AAGAAACACGAGGCTACCTGGTGGGAGGCTATCCCACTGGAGCTTCTTCTCCAGCAA 573
Db 481 AAGAAACACGAGGCTACCTGGTGGGAGGCTATCCCACTGGAGCTTCTTCTCCAGCAA 540

QY 574 AACTTGAACTAGACACTACAGCGGAGACCCACCCGGTGTGAGCGGAGCGAGGC 633
Db 541 AACTTGAACTAGACACTACAGCGGAGACCCACCCGGTGTGAGCGGAGCGAGGC 600

QY 634 CCA 636
Db 601 GCA 603

RESULT 6
BE563592
LOCUS 810 bp mRNA EST 15-AUG-2000
DEFINITION 601.4706F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3688825 5', mRNA sequence.
ACCESSION BE563592
VERSION BE563592.1 GI:9807312
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 810)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Place: LLM382 row: k column: 02
High quality sequence stop: 692.

FEATURES source

1..810
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3688825"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 176 a 284 c 205 g 145 t
ORIGIN

Query Match 47.9%; Score 588; DB 109; Length 810;
Best Local Similarity 92.1%; Pred. No. 3e-139;
Matches 753; Conservative 0; Mismatches 50; Indels 15; Gaps 12;

QY 1 ATGTGTCTACTCTCGAGCTGCGGAGCCGCGGCGGCTTCCGCTTCTGAGATCTTACCTTCGACCT 120
Db 1 ATGTGTCTACTCTCGAGCTGCGGAGCCGCGGCGGCTTCCGCTTCTGAGATCTTACCTTCGACCT 116

QY 61 CCCCTCCACCATCCCGGAGCCCGGCGGCGGCTTCCGCTTCTGAGATCTTACCTTCGACCT 180
Db 58 CCCCTCCACCATCCCGGAGCCCGGCGGCGGCTTCCGCTTCTGAGATCTTACCTTCGACCT 176

QY 121 CTCCCGGAGCCCGGAGCCCGGCGGCGGCTTCCGCTTCTGAGATCTTACCTTCGACCT 240
Db 177 CGCAGCCGAGAGGTTCTTACCTTCGAGTGTCCGCGGCCAGTCCGAGTCGAGAACCG 236

QY 241 AACCCAGCCAAAGGCTTCTTCTTCTGCTCACCACCTGCTTCTGCGCAGATCCTGATG 300
Db 237 AACCCAGCCAAAGGCTTCTTCTTCTGCTCACCACCTGCTTCTGCGCAGATCCTGATG 296

QY 301 GCTGAAGAGGTTGTCCGCGGCCCTGCTCCAGAGGAGCCCTTAACGCGGCATCCC-T 359
Db 297 GCTGAAGAGGTTGTCCGCGGCCCTGCTCCAGAGGAGCCCTTAACGCGGCATCCCCT 356

QY 360 GCGGCCACCCCTGTGTCCCGGCTTCGAGCCCTTTAATCTGACTTTCGAGGCCCTCGGA 419
Db 357 GCGGCCACCCCTGTGTCCCGGCTTCGAGCCCTTTAATCTGACTTTCGAGGCCCTCGGA 416

QY 420 CTACGCTCTGAGCTCAGCACTTTCCTCCAGACACACCCCGGCCCTTCTTAACCTGTGACT 479
Db 417 CTAGGCTCTGAGCTCAGCACTTTCCTCCAGACACACCCCGGCCCTTCTTAACCTGTGACT 474

QY 480 CCCCGCACTCCCCAAAAGAAATCCGAAAAACCAAGAAACACCAAGGCGTACCTGTGC 539
Db 475 CCCCGCACTCCCCAAAAGAAATCCGAAAAACCAAGAAACACCAAGGCGTACCTGTGC 534

```

QY 540 GCAGAGCGTATCCCACTGG-GACTTCGAGGCAACTTGAACCTCAGAAC-ACTACAGC 597
DB 535 GCAGAGCGTATCCCACTGGAGACTTCGAGGCAACTTGAACCTCAGAACACTACAGC 594
QY 598 GGAGAGCGCCACCC--GGTGGCTTCAGGCGGACCGAGCGCA-CAGAGACCGAGG-CGCAT 653
DB 595 GGAGAGCGCCAACTGGTGGCTTGTAGGCGGACCGAGCGGCA-CAGAGACCGAGGCGCAT 654
QY 654 AGAGACCGAGGCAACGCCAGCTGGGCTAGGCGGCTGGGAAGGAGAGCGCTGTTAAAT 713
DB 655 AGAGACCGAGGCAAG-CAGATGGGCTAGGCGGCTGGGAAGGAGAGCGCTCCGCAAT 713
QY 714 TATTTCTATTGCTTAATATATATATATATATATATATATATATATATATATATATAT 773
DB 714 TAATATCTTATGCTTCCTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 772
QY 774 GAGATGCTACGTAATATATATATATATATATATATATATATATATATATATATAT 811
DB 773 GACAAATGTACGTAATATATATATATATATATATATATATATATATATATATATAT 810

RESULT 7
BE910077 585 bp mRNA EST 29-SEP-2000
LOCUS BE910077 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3903754 5',
DEFINITION mRNA sequence.
ACCESSION BE910077
VERSION BE910077.1 GI:10406309
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 585)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM695 row: f column: 11
High quality sequence stop: 585.
FEATURES
Location/Qualifiers
1..585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3903754"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 131 a 119 c 188 g 147 t
ORIGIN
source

Query Match 47.6%; Score 585; DB 137; Length 585;
Best Local Similarity 100.0%; Pred. No. 1.7e-138;
Matches 585; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 544 GAGCGTATCCCACTGGGACTTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGAC 603
DB 1 GAGCGTATCCCACTGGGACTTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGAC 60

```

```

QY 604 GCCACCCGGTGTGAGGCGGGACCCAGCGGCACAGAGACCGGCGCATAGAGACCGAG 663
DB 61 GCCACCCGGTGTGAGGCGGGACCCAGCGGCACAGAGACCGGCGCATAGAGACCGAG 120
QY 664 GCACAGCCAGCTGGGGCTAGGCGGCTGGGAAGGAGAGCGCTGTTAAATTTATTTCTTAT 723
DB 121 GCACAGCCAGCTGGGGCTAGGCGGCTGGGAAGGAGAGCGCTGTTAAATTTATTTCTTAT 180
QY 724 TGTCTCTAAATTAATATATATATATATATATATATATATATATATATATATATAT 783
DB 181 TGTCTCTAAATTAATATATATATATATATATATATATATATATATATATATATAT 240
QY 784 CGTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 843
DB 241 CGTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 300
QY 844 TCTCTTGGTATTTATTTAGCTTTGTGGACTGGTGAAGCAGGACACCTGGAACCTCGGC 903
DB 301 TCTCTTGGTATTTATTTAGCTTTGTGGACTGGTGAAGCAGGACACCTGGAACCTCGGC 360
QY 904 AAAGTAGGAGAGAAATGGGGAGGACTCGGGCTGGGGAGGACGTCGCCGGCTGGGATCAAG 963
DB 361 AAAGTAGGAGAGAAATGGGGAGGACTCGGGCTGGGGAGGACGTCGCCGGCTGGGATCAAG 420
QY 964 TCTGGTGGTGGTCGTAAGTTTAGGAGGTGACATGTCATCTCCAGACATCTCAACTCCGTCT 1023
DB 421 TCTGGTGGTGGTCGTAAGTTTAGGAGGTGACATGTCATCTCCAGACATCTCAACTCCGTCT 480
QY 1024 GTCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCCATCT 1083
DB 481 GTCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCCATCT 540
QY 1084 TCTTGAAGTCGCTTTAGGGTGGCTCGGAGGTAGAGGTTGGGGG 1128
DB 541 TCTTGAAGTCGCTTTAGGGTGGCTCGGAGGTAGAGGTTGGGGG 585

RESULT 8
AW340002 586 bp mRNA EST 31-JAN-2000
LOCUS hc91d03.x1 Soares_NFL_T-GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2907365 3', mRNA sequence.
ACCESSION AW340002
VERSION AW340002.1 GI:6836628
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 586)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CCAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 471.
FEATURES
Location/Qualifiers
1..586
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2907365"
/clone_lib="Soares_NFL_T-GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: p77T3D-pac (Pharmacia) with
a modified polylinker; Site:1; Not I; Site:2; Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NIH, and B-cell

```

NCI_CGAP_CCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Donaldo.

BASE COUNT 155 a 187 c 110 g 134 t

Query Match 47.3%; Score 581.2; DB 89; Length 586;
Best Local Similarity 99.3%; Pred. No. 1.5e-137; Indels 0; Gaps 0;
Matches 583; Conservative 0; Mismatches 337;

Qy 621 GCGGACCGGACGACAGAGACGAGCGGCATAGACACGAGGACGAGCCAGCGTGGG 680
|||||
Db 586 CCGGACCGGACGACAGAGACGAGCGGCATAGACACGAGGACGAGCCAGCGTGGG 527

Qy 681 CTAGGCCGCGTGGGAAGGAGCGTCCTTAATTATTTCTTATTTCTTCTCTTAATATATT 740
|||||
Db 526 CTAGGCCGCGTGGGAAGGAGCGTCCTTAATTATTTCTTATTTCTTCTCTTAATATATT 467

Qy 741 TATATGTATTTATCTAGCTCTCTCTAGGTGATGAGATGTAGTAAATATTTTAA 800
|||||
Db 466 TATATGTATTTATCTAGCTCTCTCTAGGTGATGAGATGTAGTAAATATTTTAA 407

Qy 801 CTTATGCAAGGCTGAGATGTTCCCTCTCTGTAAATGCAAGTCTCTTGGTATTTATTG 860
|||||
Db 406 CTTATGCAAGGCTGAGATGTTCCCTCTCTGTAAATGCAAGTCTCTTGGTATTTATTG 347

Qy 861 AGCTTTGTGGACTGTGGGAAGCAGGACACCTTGAACCTGCGGCAAGTAGAGAAGAAAT 920
|||||
Db 346 AGCTTTGTGGACTGTGGGAAGCAGGACACCTTGAACCTGCGGCAAGTAGAGAAGAAAT 287

Qy 921 GCGGAGACTCGGCTGGGGGAGGACGTCCTCGGTGGGATGAAGTCTGGTGGTGGTGGT 980
|||||
Db 286 GGGGAGACTCGGCTGGGGGAGGACGTCCTCGGTGGGATGAAGTCTGGTGGTGGTGGT 227

Qy 981 AGTTAGAGGTGACTGCATCTCCAGCATCTCAACTCCGCTCTCTACTGTGTGAGACT 1040
|||||
Db 226 AGTTAGAGGTGACTGCATCTCCAGCATCTCAACTCCGCTCTCTACTGTGTGAGACT 167

Qy 1041 TCGGCGGACCATTAGGAATGAGATCCGTGAGATCTTCCATCTCTTGAAGTCCCTTTA 1100
|||||
Db 166 TCGGCGGACCATTAGGAATGAGATCCGTGAGATCTTCCATCTCTTGAAGTCCCTTTA 107

Qy 1101 GCGTGGCTGCCAGGTAGAGGCTTGGGGTTCGTCGCGCTGTACGAGCGACTGTGAGAT 1160
|||||
Db 106 GCGTGGCTGCCAGGTAGAGGCTTGGGGTTCGTCGCGCTGTACGAGCGACTGTGAGAT 47

Qy 1161 GCGCTAGTATGTTCTGTGAACACAAATAAAATGATTACTGTCAA 1206
|||||
Db 46 GCGCTAGTATGTTCTGTGAACACAAATAAAATGATTACTGTCAA 1

RESULT 9
AM662348/c 598 bp mRNA EST 06-APR-2000
LOCUS hi25c07.x1 NCI_CGAP_Co14 Homo sapiens cdna clone IMAGE:2973324 3',
DEFINITION mRNA sequence.

ACCESSION AM662348
VERSION AM662348.1 GI:7454886
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 598)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -40UP from Gibco
High quality sequence stop: 392.
Location/Qualifiers
1..598
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2973324"
/clone_lib="NCI_CGAP_Co14"
/issue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: Sall;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

FEATURES
source

BASE COUNT 163 a 187 c 116 g 132 t

Query Match 47.2%; Score 579.6; DB 93; Length 598;
Best Local Similarity 98.5%; Pred. No. 3.9e-137;
Matches 585; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 611 GGTGCTTGAGCGGGACCGAGCGCCACAGAGCCGAGGCGCATAGAGCCGAGGCACAGC 670
|||||
Db 598 GGTGCTTGAGCGGGACCGAGCGCCACAGAGCCGAGGCGCATATAGCCGAGGCACAGC 539

Qy 671 CGAGCTGGGCTAGGCGCGGTGGGAAGGAGAGCGTGGTAAATTTATTTCTTATTGCTCT 730
|||||
Db 538 CGAGCTGGGCTTAGGCGCGGTGGGAAGGAGAGCGTGGTAAATTTATTTCTTATTGCTCT 479

Qy 731 AATTAATATTTATATGTATTTATGTACCTCTCTAGGTGATGAGATGTACGTAATA 790
|||||
Db 478 AATTAATATTTATATGTATTTATGTACCTCTCTAGGTGATGAGATGTACGTAATA 419

Qy 791 TTTATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTCTGCTCTGCTCTGCTCT 850
|||||
Db 418 TTTATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTCTGCTCTGCTCTGCTCT 359

Qy 851 GTATTTATTCAGCTTTGTGGAGTGGTGGGAAGCAGACACCTGGAACCTGCGCAAGTAG 910
|||||
Db 358 GTATTTATTCAGCTTTGTGGAGTGGTGGGAAGCAGACACCTGGAACCTGCGCAAGTAG 299

Qy 911 GAGAAGAAATGGGAGGACTCGGCTGGGGGAGGACCTCCGCGCTGGGATCAAGTCTGGTG 970
|||||
Db 298 GAGAAGAAATGGGAGGACTCGGCTGGGGGAGGACCTCCGCGCTGGGATCAAGTCTGGTG 239

Qy 971 GTGGGTGTAAGTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCGCTGCTCTACT 1030
|||||
Db 238 GTGGGTGTAAGTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCGCTGCTCTACT 179

Qy 1031 GTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCCTTCCATCTCTTGAA 1090
|||||
Db 178 GTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCCTTCCATCTCTTGAA 119

Qy 1091 GTCGCCTTTAGGTGGCTGCGAGGTAGAGGGTTGGGGTGGTGGGCTGTACCGAGCGA 1150
|||||
Db 118 GTCGCCTTTAGGTGGCTGCGAGGTAGAGGGTTGGGGTGGTGGGCTGTACCGAGCGA 59

Qy 1151 CTGTCGAGATCGCTAGTATGTTCTGTGAACACAAATAAAATGATTACTGTGTC 1204
|||||
Db 58 CTGTCGAGATCGCTAGTATGTTCTGTGAACACAAATAAAATGATTACTGTGTC 5

```

RESULT 10
BE621600/c
LOCUS
DEFINITION
601493736T1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3895736 3',
mRNA sequence.
ACCESSION
BE621600
VERSION
BE621600.1 GI:9892540
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 614)
AUTHORS
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue procurement: ATCC
cDNA library preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM674 row: h column: 09
High quality sequence start: 12
High quality sequence stop: 613.
FEATURES
source
1..614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3895736"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: Not1;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT
158 a 198 c 123 g 135 t
ORIGIN

Query Match
Best Local Similarity 98.2%; Score 567.2; DB 110; Length 614;
Matches 595; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

Qy 582 CTCAGACACTACAGCGGACAGCCACCCGGTCTTTCAGCGGACCCAGCGGCACACAGAG 641.
|||||
Db 614 CTCAGACACTACAGCGGACAGCCACCCGGTCTTTCAGCGGACCCAGCGGCACACAGAG 555
|||||
Qy 642 ACCGAGCGCATAGACAGCCAGGACAGCCAGCTGGGGGTAGGCCCGGTGGGAAGGAGA 701
|||||
Db 554 ACCGAGCGCATAGACAGCCAGGACAGCCAGCTGGGGGTAGGCCCGGTGGGAAGGAGA 495
|||||
Qy 702 GCGTCGTGTA--TTTATTCTTATTGCTCTTAATATATATATATATATATATATATGTCGT 759
|||||
Db 494 GCGTCGTGTAATGTTATGTTCTTATTGCTCTTAATATATATATATATATATATATGTCAGT 435
|||||
Qy 760 CTCTCTAGGTCATGATGTCAGTGAATATATATATATATATATATATATATATATATATATAT 819
|||||
Db 434 CTCTCTAGGTCATGATGTCAGTGAATATATATATATATATATATATATATATATATATATAT 375
|||||
Qy 820 TGTTCCTCTGCTTAATGACAGTCTCTTGGTATTTATTGAGCTTTTGTGGAGCTGGTGG 879
|||||
Db 374 TGTTCCTCTGCTTAATGACAGTCTCTTGGTATTTATTGAGCTTTTGTGGAGCTGGTGG 315
|||||
Qy 880 AAGCAGACACCTCGAGCTCGGGCAAGTAGAGAGAGAAATGUGGAGGA-CTCGGGTGGG 938
|||||

```

```

Db 314 AAGCAGGACACCTGGAACTCGGCAAGTAGGAGAGAAATGGGAGGATCTCGSGTGGG 255
Qy 939 GGAGGACGTCCTCCGGCTGGGATGAAGTCTGGTGGTGGTAACTTTAGGAGGTCACATGC 998
|||||
Db 254 GGAGGACGTCCTCCGGCTGGGATGAAGTCTGGTGGTGGTAACTTTAGGAGGTCACATGC 195
Qy 999 ATCTCCAGCATCTCAACTCCGCTCTGCTACTGTGTGAGACTTCGGCGGACCATAGGAA 1058
|||||
Db 194 ATCTCCAGCATCTCAACTCCGCTCTGCTACTGTGTGAGACTTCGGCGGACCATAGGAA 135
Qy 1059 TGAGATCCGTGAGATCCCTTCCATCTTGAAGTCGCTTTAGGGTGGCTCGGAGGTAGA 1118
|||||
Db 134 TGAGATCCGTGAGATCCCTTCCATCTTGAAGTCGCTTTAGGGTGGCTCGGAGGTAGA 75
Qy 1119 GGGTGGGGTGGTGGGCTGTACAGGAGGACACTGTGAGATCCGCTAGTATCTTCTCTG 1178
|||||
Db 74 GGGTGGGGTGGTGGGCTGTACAGGAGGACACTGTGAGATCCGCTAGTATCTTCTCTG 15
Qy 1179 AACACA 1184
|||||
Db 14 CACAA 9
|||||

RESULT 11
AW956285
LOCUS
DEFINITION
EST368355 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
ACCESSION
AW956285
VERSION
AW956285.1 GI:8145968
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 588)
AUTHORS
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holl
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL
Unpublished (2000)
COMMENT
Contact: John Quackenbush
The Institute for Genomic
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 90
Seq primer: Reverse.
Location/Qualifiers
1..588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGD"
/note="Vector: pBluescriptSkm"
BASE COUNT
135 a 117 c 187 g 149 t
ORIGIN

Query Match
Best Local Similarity 98.3%; Score 561.6; DB 97; Length 588;
Matches 578; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 556 AACTGGGCACTTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGACGCCACCGTGC 615
|||||
Db 1 AACTGGGCACTTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGACGCCACCGTGC 60
|||||
Qy 616 TTGAGCGGGACCCGAGCGGACAGACCGAGCGGCGCATAGAGACCGAGGACACAGCCGAGC 675
|||||
Db 61 TTGAGCGGGACCCGAGCGGACAGACCGAGCGGCGCATAGAGACCGAGGACACAGCCGAGC 120
|||||
Qy 676 TGGGGCTAGGCCCGGTGGGAAGGAGACGCTGTTAATTTATTTCTTATTTCTCTAATTA 735
|||||

```

```
Db 121 TGGGCTAGCCGCGTGGGAAGGAGAGCGTCTTAATTTATTTCTTATTTGCTCCCTAATTA 180
Qy 736 ATATTTATATGTTATTTATGATGCTCTCTAGGTGATGAGATGCTAGTAAATATTTAT 795
Db 181 ATATTTATATGTTATTTATGATGCTCTCTAGGTGATGAGATGCTAGTAAATATTTAT 240
Qy 796 TTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTCTGTTAAATGAGGCTCTTGGTAT 855
Db 241 TTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTCTGTTAAATGAGGCTCTTGGTAT 300
Qy 856 TATTGAGCTTTGTGGAGCTGTGGAGAGGAGGACCTTGGAACTGCGGCAAAAGTAGGAA 915
Db 301 TATTGAGCTTTGTGGAGCTGTGGAGAGGAGGACCTTGGAACTGCGGCAAAAGTAGGAA 360
Qy 916 GAAATGGGAGGACTCGGGTGGGAGGAGGAGCTCCGGGTGGAGTAACTCTGGTGGTGG 975
Db 361 GAAATGGGAGGACTCGGGTGGGAGGAGGAGCTCCGGGTGGAGTAACTCTGGTGGTGG 420
Qy 976 TCGTAAGTTTGGAGGCTGACTGCTATCCCTCCAGCATCTCAACTCCGCTCTGCTACTGCTG 1035
Db 421 TCGTAAGTTTGGAGGCTGACTGCTATCCCTCCAGCATCTCAACTCCGCTCTGCTACTGCTG 480
Qy 1036 AGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCCTTCCATCTCTTGAAGTCG 1095
Db 481 AGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCCTTCCATCTCTTGAAGTCG 540
Qy 1096 CTTTAGGCTGGCTCGAGGT-AGAGGGTTGGGGTGGGGTGGTGGTGGTGGTGA 1142
Db 541 CTTTAGGCTGGCTCGAGGTAAAAGGGTTGGGGTGGGGTGGTGGTGGTGA 588

RESULT 12
AV690660 617 bp mRNA EST 25-SEP-2000
LOCUS AV690660 GK Homo sapiens cDNA clone GKCBA04 5', mRNA sequence.
DEFINITION AV690660
VERSION AV690660.1 GI:10292523
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Wu, T., Qian, B., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H.,
X., X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L.,
X., S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang
Y., Gu, Y., Chen, Z., and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@hgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1. 617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBA04"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
131 a 182 c 164 g 139 t 1 others
BASE COUNT
ORIGIN

Query Match 45.4%; Score 557; DB 38; Length 617;
```

Best Local Similarity 95.5%; Pred. No. 2.3e-131;
Matches 596; Conservative 0; Mismatches 21; Indels 7; Gaps 2;

```
Qy 277 ATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCCCTGCTCCAGAG 336
Db 1 ATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCCCTGCTCCAGAG 60
Qy 337 GACGCCCTTAACGGCGCATCCTGCGGCCACCCCTGTGTCGCCCGTCTCGAGCCCTTT 396
Db 61 GACGCCCTTAACGGCGCATCCTGCGGCCACCCCTGTGTCGCCCGTCTCGAGCCCTTT 120
Qy 397 AATCTGACTTCGAGCCCTCGGACTTACGCTCTGACCTTCAGACTTTCCTCCAGCAAC 456
Db 121 AATCTGACTTCGAGCCCTCGGACTTACGCTCTGACCTTCAGACTTTCCTTCAGCAAC 180
Qy 457 CCGGCCCTCTTAACTGTGACTCCCGGACTCCCGCAAAAGATCGCAAAACACAAA 516
Db 181 CCGGCCCTCTTAACTGTGACTTCCCGGACTTCCCAAAAAGATCGC---ACCACAAA 236
Qy 517 GAAACACGAGCGTACCTGGTGCAGAGAGCTATCCCAACTGGGACTTCCGAGGCAAC 576
Db 237 GAAACACGAGCGTACCTGGTGCAGAGAGCTATCCCAACTGGGACTTCCGAGGCAAC 296
Qy 577 TTGAATCAGAACACTACAGCGGAGAGCCGACCCGGTCTTGAGCGGGGACCGAGCGCA 636
Db 297 TTGAATCAGAACACTACAGCGGAGAGCCGACCCGGTCTTGAGCGGGGACCGAGCGCA 356
Qy 637 CAGAGACGAGGCGCATAGACCGAGGCGGACCGGCTAGGGCTAGGGCCGCTGGGAA 696
Db 357 CAGAGACGAGGCGCATAGACCGGCGGACCGGCTAGGGCTAGGGCCGCTGGGAA 416
Qy 697 GGAGAGCGTCTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 756
Db 417 GGAGAGCGTCTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 476
Qy 757 CGTCTCTAGGTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 816
Db 477 CGTCTCTAGGTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 536
Qy 817 AGATGTTCCCTCTGTTAAATGAGTCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 876
Db 537 AGATGTTCCCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1
Qy 877 TGAAGCAGGACGACCTTGGAACTGC 900
Db 594 GGAAGCAGGACGACCTTGGCTGCGGC 617
```

RESULT 13

AW263252/c

LOCUS

DEFINITION

AW263252

AW263252

AW263252.1

GI:6640068

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 548)

NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 466.

Location/Qualifiers


```
source          1. 548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2700722"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pMT73D-Pac (pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell.
NCL CGAP.GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT    157 a 171 c 98 g 122 t
ORIGIN

Query Match      44.4%; Score 544.8; DB 88; Length 548;
Best Local Similarity 99.6%; Pred. No. 2.8e-128;
Matches 546; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 660 CGAGGCACACCCAGCTGGGCTAGGCCCGGTGGGAGGAGAGCGTCGTTAATTATTC 719
|||||
DB 548 CGAGGCACACCCAGCTGGGCTAGGCCCGGTGGGAGGAGAGCGTCGTTAATTATTC 489

QY 720 TTATTCCTCTAATTAATATTTATATGATATTTATGACGTCCTCCAGTGAGATG 779
|||||
DB 488 TTATTCCTCTAATTAATATTTATGATATTTATGACGTCCTCCAGTGAGATG 429

QY 780 TGACGTAAATATTTATTTAACTATGCAAGGCTGAGATGTTCCCTCTGCTTAATG 839
|||||
DB 428 TGACGTAAATATTTATTTAACTATGCAAGGCTGAGATGTTCCCTCTGCTTAATG 369

QY 840 CAGGTCCTCTGGTATTTATTGAGCTTTGTGGGACTGGTGGAGGAGACACCTGGAACTG 899
|||||
DB 368 CAGGTCCTCTGGTATTTATTGAGCTTTGTGGGACTGGTGGAGGAGACACCTGGAACTG 309

QY 900 CGCAAAAGTAGAGACAAATAGGGAGGACTCGGGTGGGGGAGGAGACTCCCGGCTGGAT 959
|||||
DB 308 CGCAAAAGTAGAGACAAATAGGGAGGACTCGGGTGGGGGAGGAGACTCCCGGCTGGAT 249

QY 960 GAAGTCTGGTGGTGGTGGTAAAGTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCC 1019
|||||
DB 248 GAAGTCTGGTGGTGGTGGTAAAGTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCC 189

QY 1020 GTCTGTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCTCTTC 1079
|||||
DB 188 GTCTGTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCTCTTC 129

QY 1080 ATCTCTTGAAGTCGCCCTTTAGGGTGGCTCCGAGGTAGAGGGTTGGGGGTTGGGGCTG 1139
|||||
DB 128 ATCTCTTGAAGTCGCCCTTTAGGGTGGCTCCGAGGTAGAGGGTTGGGGGTTGGGGCTG 69

QY 1140 TCAGGAGGACACTGTCGAGATCGCTAGTATGTTCTGTGAACAAATAAATGATTTA 1199
|||||
DB 68 TCAGGAGGACACTGTCGAGATCGCTAGTATGTTCTGTGAACAAATAAATGATTTA 9

QY 1200 CTGTCAAA 1207
|||||
DB 8 CTGTCAAA 1
```

```
RESULT 14
BE878459
LOCUS
DEFINITION BE878459 1070 bp mRNA EST 27-SEP-2000
mRNA sequence.
ACCESSION BE878459
```

```
VERSION BE878459.1 GI:10327235
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1070)
AUTHORS NIH-MGC Instt://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DTB/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM660 row: c column: 17
High quality sequence stop: 645.
```

FEATURES

```
Source
1..1070
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3890248"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: Not 1;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
```

```
BASE COUNT 276 a 324 c 302 g 168 t
ORIGIN
```

```
Query Match      44.0%; Score 540.8; DB 137; Length 1070;
Best Local Similarity 95.6%; Pred. No. 3.3e-127;
Matches 675; Conservative 0; Mismatches 17; Indels 14; Gaps 11;

QY 151 CGCCCCAGGCCCTCTCCGGGACCGAAAGCGAGCGG-CAGGGTTCTCTACCTCGA-C 208
DB 1 CGCCCCAGGCCCTCTCCGGGACCGAAAGCGAGCGGTCACGGTTCTCTACCTCGATG 60

QY 209 TGTCTCGGCGCCAG-CTGCCAGTCGAGGACCGAACCCAGCCAA-RAGGCTTCTCTTCT 266
|||||
DB 61 TGTCTCGGCGCCAGTCTGCCAGTCGAGGAACCGAACCCAGCCAAAGAGGTTCTCTTCT 120

QY 267 GCTGC-TCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGCGCCCC 325
|||||
DB 121 GCTGCATCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCC 180

QY 326 TGCTCCAGAGGAC-GCCCTTAACGCCGATCCTCTGGC-CCCAACCCCTGTGTCGCCGT 383
|||||
DB 181 TGCTCCAGAGGACTGCCCTTAACGCCGATCCTCTGGCTGCCACCCCTGTGTCGCCGT 240

QY 384 CCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGACCTCAGCCTTT 443
|||||
DB 241 CCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGACCTCAGCCTTT 300

QY 444 CCTCCAGAACACCCCGCCCTTCTAACTGTGACTTCCCGGCACTCCCAAAAGAAATCC 503
|||||
DB 301 CCTCCAGAACACCCCGCCCTTCTAACTGTGACTTCCCGGCACTCCCAAAAGAAATCC 360

QY 504 GAAAAACCAAGAAACACACAGCGCTACTGCTGGCGGAGAGCGTATCCCCAACCTGGGA 563
|||||
DB 361 GAAAAACCAAGAAACACACAGCGCTACTGCTGGCGGAGAGCGTATCCCCAACCTGGGA 420

QY 564 CTTCCGAGGCACTTGAACCTCAGAACACTACACGCGGAGCGCCACCTGGTGTGAGGCG 623
|||||
DB 421 CTTCCGAGGCACTTGAACCTCAGAACACTACACGCGGAGCGCCACCTGGTGTGAGGCG 480
```


QY 624 GGACCGAGGCGACAGAGACCGAGGCGCATAGAGACCGAGGACAGCCAGCTGGGCGTA 683
|||||
Db 481 GGACCGAGGCGACAGAGACCGAGGCGCATAGAGACCGAGGACAGCCAGCTGGGCGTA 540
|||||
QY 684 GGCCCGGTGGG-AAGGAGAGCGTGGTAAATTTATTTCTTATTTGCTCTCCTAATTAATTTA 742
|||||
Db 541 GGCCCGGTGGGAAAGAGAGCGTCCGTAATTTATTTCTTATTTGCTCTCCTAATTAATTA 599
|||||
QY 743 TATGTAATTTATGACGTCCCTCCTAGG----TGATGGAGATGCTGACGTAATTAATTTT 798
|||||
Db 600 TATGTAATTTATGACGTCCCTCCTAGGCTGATGACGACTGTGACCGCTAATTAATTTAT 659
|||||
QY 799 AACTTATGCAAGGCTGAGATGTCCTCT-GCTGTAATGACAGG 843
|||||
Db 660 AACTTATGCAAGGCTGAGATGTCCTCTGCTGCTGTAATGACAGG 705
|||||

RESULT 15
AI800828/c
LOCUS
DEFINITION
wgl3f12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2365007 3', mRNA sequence.
ACCESSION
AI800828
VERSION
AI800828.1 GI:5366300
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 521)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 662 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1..521
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2365007"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pMT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 154 a 159 c 88 g 120 t
ORIGIN

Query Match 42.3%; Score 519.4; DB 25; Length 521;
Best Local Similarity 99.8%; Pred. No. 8.3e-122;
Matches 520; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 690 GTGGAGGAGAGCGGTCGTTAAATTTATTTCTTATTTGCTCTCCTAATTAATTTATATATAT 749
|||||
Db 521 GTGGAGGAGAGCGGTCGTTAAATTTATTTCTTATTTGCTCTCCTAATTAATTTATATATAT 462
|||||
QY 750 TTATGTACGTCCCTCCTAGGTGATGGAGATGTGTACGTAATTAATTTATTTTAACTTATGCAA 809
|||||
Db 461 TTATGTACGTCCCTCCTAGGTGATGGAGATGTGTACGTAATTAATTTTAACTTATGCAA 402
|||||
QY 810 GGTGTGAGATGTTCCCTCTGCTGTAATGCAGGTCTCTGGTAATTTATTTGAGCTTTGTG 869
|||||
Db 401 GGTGTGAGATGTTCCCTCTGCTGTAATGCAGGTCTCTGGTAATTTATTTGAGCTTTGTG 342
|||||
QY 870 GGAAGTGTGGAAGCAGGACACCTGGAACCTGGGCAAGTAGGAGAGAAATGGGAGGAC 929
|||||
Db 341 GGAAGTGTGGAAGCAGGACACCTGGAACCTGGGCAAGTAGGAGAGAAATGGGAGGAC 282
|||||
QY 930 TCGGGTGGGAGGAGCGTCCCGGCTGGGATGAAGTCTGGTGGTGGTCTGTAAGTTAGGA 989
|||||
Db 281 TCGGGTGGGAGGAGCGTCCCGGCTGGGATGAAGTCTGGTGGTGGTCTGTAAGTTAGGA 222
|||||
QY 990 GGTGACTGTCATCTCCAGCATCTCAACTCCGCTCTGCTACTGTGTGAGACTTCGGCGGAC 1049
|||||
Db 221 GGTGACTGTCATCTCCAGCATCTCAACTCCGCTCTGCTACTGTGTGAGACTTCGGCGGAC 162
|||||
QY 1050 CATTAGGAATGAGATCCGTGAGATCCTTCCATCTTCTGAAGTCGCGCTTTAGGTTGGCTG 1109
|||||
Db 161 CATTAGGAATGAGATCCGTGAGATCCTTCCATCTTCTGAAGTCGCGCTTTAGGTTGGCTG 102
|||||
QY 1110 CGAGGTAGAGGGTTGGGGGTTGGTGGGCTGTACGAGCGGACTGTGAGATCGCCTAGTA 1169
|||||
Db 101 CGAGGTAGAGGGTTGGGGGTTGGTGGGCTGTACGAGCGGACTGTGAGATCGCCTAGTA 42
|||||
QY 1170 TGTCTGTGAACACAAATAAAATTTGATTTACTCTCAAAAAA 1210
|||||
Db 41 TGTCTGTGAACACAAATAAAATTTGATTTACTCTCAAAAAA 1

Search completed: January 30, 2001, 18:47:25
Job time: 28456 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 22:04:48 ; Search time 2136.76 Seconds
(without alignments)
509,900 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468
Perfect score: 258
Sequence: 1 GTCCGGCGCAGTCCAGT.....AGCAACACCCGGCGCCTTC 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 12325308 seqs, 2111498085 residues

Total number of hits satisfying chosen parameters: 24650616

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/pna/pctus_COMB.seq.*
 - 2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
 - 3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
 - 4: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
 - 6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
 - 7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
 - 8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
 - 9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
 - 10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
 - 11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
 - 12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
 - 13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
 - 14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
 - 15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
 - 16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
 - 17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
 - 18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
 - 19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
 - 20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
 - 21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
 - 22: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
 - 23: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
 - 24: /cgn2_6/ptodata/2/pna/US097_COMB.seq.*
 - 25: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
 - 26: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
 - 27: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
 - 28: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
 - 29: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
 - 30: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
 - 31: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
 - 32: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
 - 33: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
 - 34: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
 - 35: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
 - 36: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
 - 37: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
 - 38: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
 - 39: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
 - 40: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
 - 41: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
 - 42: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
 - 43: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	258	100.0	453	19	US-09-528-409-88514
2	258	100.0	506	24	US-09-726-805-429
3	258	100.0	521	22	US-09-644-869-356
4	258	100.0	521	44	US-60-196-718-2355
5	258	100.0	529	23	US-09-699-998-4785
6	258	100.0	617	17	US-09-396-885-430
7	258	100.0	657	19	US-09-522-303-1025
8	258	100.0	673	23	US-09-698-010-14935
9	258	100.0	673	23	US-09-699-998-9797
10	258	100.0	673	24	US-09-710-286-3772
11	258	100.0	673	24	US-09-716-473-2500
12	258	100.0	673	24	US-09-721-588-4548
13	258	100.0	673	55	US-09-726-787-2982
14	258	100.0	706	23	US-09-699-998-1778
15	258	100.0	803	23	US-09-699-998-10469
16	258	100.0	888	23	US-09-652-109-10087
17	258	100.0	888	23	US-09-652-126-9303
18	258	100.0	888	23	US-09-652-816-9065
19	258	100.0	888	24	US-09-716-473-3223
20	258	100.0	888	24	US-09-716-990-1375
21	258	100.0	888	24	US-09-721-588-4181
22	258	100.0	888	24	US-09-726-172-2642
23	258	100.0	888	24	US-09-726-805-1977
24	258	100.0	888	55	US-09-726-805-2896
25	258	100.0	1228	11	US-08-799-910-9
26	258	100.0	1228	12	US-08-825-486-9
27	258	100.0	1228	12	US-08-826-248-9
28	258	100.0	1228	12	US-08-870-434-5
29	258	100.0	1228	13	US-08-925-588-9
30	258	100.0	1228	17	US-09-372-044-9
31	258	100.0	1308	1	PCT-US00-15136-46
32	258	100.0	1316	17	US-09-338-425-2393
33	258	100.0	1316	22	US-09-649-162-9722
34	258	100.0	1316	23	US-09-652-109-9454
35	258	100.0	1316	23	US-09-652-128-10214
36	258	100.0	1316	23	US-09-652-914-9119
37	258	100.0	1316	23	US-09-699-998-9799
38	258	100.0	1316	24	US-09-710-280-2658
39	258	100.0	1316	24	US-09-716-473-2888
40	258	100.0	1316	24	US-09-721-588-3900
41	258	100.0	1316	24	US-09-726-171-2410
42	258	100.0	1316	24	US-09-726-790-2393
43	258	100.0	1316	24	US-09-726-791-1703
44	258	100.0	1316	24	US-09-726-805-1978
45	258	100.0	1316	55	US-09-726-787-2984

ALIGNMENTS

RESULT 1

US-09-528-409-88514
; Sequence 88514, Application US/09528409
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labele, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark
; APPLICANT: Jones, Lee W.
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; FILE REFERENCE: 774
; CURRENT APPLICATION NUMBER: US/09/528.409
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,453
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 116231
; SOFTWARE: Hy-patent.pl Version 3.1
; SEQ ID NO 88514
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(453)
; OTHER INFORMATION: n = A,T,C or G

US-09-528-409-88514

Query Match 100.0%; Score 258; DB 19; Length 453;
Best Local Similarity 100.0%; Pred. No. 7.4e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCGAACCCAGCAAGAGCTTCTCTTTCTGCTG 60
DB 51 gtcggcgccagctgccagtcgaggaacccagcaagagcttctcttctgctg 110
QY 61 CTCACATGCTTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTGCCT 120
DB 111 ctcaccatgcttcttctgccagatcctgagtgagaggtgtgcccggccctgcct 170
QY 121 CCAGAGAGCCCTTACGCGGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAG 180
DB 171 ccagagagcccttaacgcgcgcatccctggcgccccaccctgtgtcccccgcctcag 230
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACCTTCTCCAG 240
DB 231 ccctttaatctgacttcggagccctcggactacgctcgtgacctcagcacttctccag 290
QY 241 CAACACCCGCGCCCTTC 258
DB 291 caacacccgcgcccttc 308

RESULT 2

US-09-726-805-429
; Sequence 429, Application US/09726805
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.2017-001
; CURRENT APPLICATION NUMBER: US/09/726.805
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/168,140
; NUMBER OF SEQ ID NOS: 2158
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-726-805-429

Query Match 100.0%; Score 258; DB 24; Length 506;
Best Local Similarity 100.0%; Pred. No. 7.5e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCGAACCCAGCAAGAGCTTCTCTTTCTGCTG 60
DB 138 gtcggcgccagctgccagtcgaggaacccagcaagagcttctcttctgctg 197
QY 61 CTCACATGCTTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTGCCT 120
DB 198 ctcaccatgcttcttctgccagatcctgagtgagaggtgtgcccggccctgcct 257
QY 121 CCAGAGAGCCCTTACGCGGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAG 180
DB 258 ccagagagcccttaacgcgcgcatccctggcgccccaccctgtgtcccccgcctcag 317
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACCTTCTCCAG 240
DB 318 ccctttaatctgacttcggagccctcggactacgctcgtgacctcagcacttctccag 377
QY 241 CAACACCCGCGCCCTTC 258
DB 378 caacacccgcgcccttc 395

RESULT 3

US-09-644-869-356
; Sequence 356, Application US/09644869
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Monehan, John
; APPLICANT: Richardson, Jennifer
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.1182-001
; CURRENT APPLICATION NUMBER: US/09/644.869
; PRIOR FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: 60/151,062
; NUMBER OF SEQ ID NOS: 9708
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 356
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-644-869-356

Query Match 100.0%; Score 258; DB 22; Length 521;
Best Local Similarity 100.0%; Pred. No. 7.5e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCGAACCCAGCAAGAGCTTCTCTTTCTGCTG 60
DB 211 gtcggcgccagctgccagtcgaggaacccagcaagagcttctcttctgctg 270
QY 61 CTCACATGCTTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTGCCT 120
DB 271 ctcaccatgcttcttctgccagatcctgagtgagaggtgtgcccggccctgcct 330
QY 121 CCAGAGAGCCCTTACGCGGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAG 180
DB 331 ccagagagcccttaacgcgcgcatccctggcgccccaccctgtgtcccccgcctcag 390
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACCTTCTCCAG 240
DB 391 ccctttaatctgacttcggagccctcggactacgctcgtgacctcagcacttctccag 450

```
Best Local Similarity 100.0%; Pred. No. 7.5e-52; Gaps -0;
Matches 258; Conservative 0; Mismatches 0; Indels 0;
```

1	QY	GTCCGGGCCAGCTGCCAGCTCCAGGGAACCGAACCCAGCCAAAGGCTTCCTTCTGCTG	80
182	Db	gtccggcgccagctgcagtcgaggaaacggaacccagcgaaggctctcttctctgctg	241
61	QY	CTCACCATCGCTCTTCTGCCAGATCCCTGATGGCTGAAGAGGTGTCGCGGCCCTTGCT	120
242	Db	ctcaccatcgctctctcgccagatcctgctgctgaaggaggtgTgcgaggccctgacct	301
121	QY	CCAGAGACGCCCTAACGCCGATCCCTGCGGCCACCCCTGTCTCCCCGTCTCCAG	180
302	Db	ccagaggacgcccctaaagcgcatacctggcgcccaacctgtgccccgtctctgag	361
181	QY	CCCTTTTAATGTGATTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG	240
362	Db	ccctttaatctgactctggagccctcgaactcgtctggacctcagcaattctctccag	421
241	QY	CAACACCGCGCGCCTTC	258
422	Db	caacacccggcgcccttc	439

RESULT 6
US-09-396-885-430

Query Match	100.0%	Score 258;	DB 17;	Length 617;
Best Local Similarity	100.0%;	Pred. No. 7.6e-52;		
Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

241	CAACACCCGGCCGCTTC	258
451	caacacccggccgcttc	468

```

RESULT      4
US-60-196-718-2355/c
; Sequence 2355, Application US/60196718
;
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO00456
; CURRENT APPLICATION NUMBER: US/60/196,718
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2355
; LENGTH: 521
; TYPE: DNA
; ORGANISM: HUMAN
;
US-60-196-718-2355

```

Query Match	100.0%	Score 258;	DB 44;	Length 521;	
Best Local Similarity	100.0%;	Pred. No. 7.5e-50;			
Matches 258;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
1	GTCCGGCGCCAGTGGCAGTCGAGGAACCCAGCCAAAGAGCTTCTCTTCTGCTG 60				
400	GTCCGGCGCCAGTGGCAGTCGAGGAACCCAGCCAAAGAGCTTCTCTTCTGCTG 341				
61	CTCACCATGCTCTCTCCAGATCCTGATGGCTGAAGAGGTGTGCGCGGCCCTGCCT 120				
340	CTCACCATGCTCTCTCCAGATCCTGATGGCTGAAGAGGTGTGCGCGGCCCTGCCT 281				
121	CCAGAGGACGCCCTTAACGGCGCATCCTCGCGCCCAACCCTGTGTGTCGCCCGCTCCGAG 180				
280	CCAGAGGACGCCCTTAACGGCGCATCCTGCGCGCCAACCCCTGTGTGTCGCCCGCTCCGAG 221				
181	CCCTTTTAATCTCACTTCGGAGCCCTTCGGACTACGCTCTGGAGCCTCAGACCTTCTCCTCCAC 240				
220	CCCTTTTAATCTCACTTCGGAGCCCTTCGGACTACGCTCTGGAGCCTCAGACCTTCTCCTCCAG 161				
241	CAACACCCGGCGCCCTTC 258				
160	CAACACCCGGCGCCCTTC 143				

```

RESULT      5
US-09-699-998-4785
; Sequence 4785, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4785
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-699-998-4785

```

Query Match

QY 241 CAACACCCGCCGCTTC 258
Db 346 caacacccggcgcccttc 363

RESULT 7

US-09-522-303-1025
; Sequence 1025, Application US/09522303
; GENERAL INFORMATION:
; APPLICANT: Holtzman, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES DERIVED FROM A
; TITLE OF INVENTION: HUMAN MAMMARY EPITHELIUM LIBRARY
; FILE REFERENCE: 1600.1086-001
; CURRENT APPLICATION NUMBER: US/09/522,303
; PRIOR FILING DATE: 2000-03-08
; EARLIER APPLICATION NUMBER: 60/123,393
; NUMBER OF SEQ ID NOS: 1353
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1025
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-522-303-1025

Query Match 100.0%; Score 258; DB 19; Length 657;
Best Local Similarity 100.0%; Pred. No. 7.6e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCCAGTCCAGGAAACCGAACCCAGCAAAAGGCTTCTCTTCTGCTG 60
Db 277 gtccggcgccagctgccagtcgaggaacccagcaaaagcttcttcttctgctg 336
QY 61 CTCACATCGTCTTCTGCGACATCTGATGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120
Db 337 ctccacatcgcttctgcccagatctctgctgaagagggtgtgccggcgccctgcct 396
QY 121 CCAGAGGACGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTCGCCGCGCTCGAG 180
Db 397 ccagaggacgcccctaacgcgcgcctccctggcgccacccctgtgtcccccgtctcgag 456
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCATTCTCTCCAG 240
Db 457 ccccttaactgacttcggagccctcgagctacgctctgagctcagcacttctctccag 516
QY 241 CAACACCCGCCGCTTC 258
Db 517 caacacccggcgcccttc 534

RESULT 8

US-09-698-010-14935
; Sequence 14935, Application US/09698010
; GENERAL INFORMATION:
; APPLICANT: Williamson, Mark
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2029-001
; CURRENT APPLICATION NUMBER: US/09/698,010
; PRIOR FILING DATE: 2000-10-27
; EARLIER APPLICATION NUMBER: 60/162,358
; NUMBER OF SEQ ID NOS: 15684
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14935
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-698-010-14935

Query Match 100.0%; Score 258; DB 23; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCCAGTCCAGGAAACCGAACCCAGCAAAAGGCTTCTCTTCTGCTG 60
Db 269 gtccggcgccagctgccagtcgaggaacccagcaaaagcttcttcttctgctg 328
QY 61 CTCACATCGTCTTCTGCGACATCTGATGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120
Db 329 ctccacatcgcttctgcccagatctctgctgaagagggtgtgccggcgccctgcct 388
QY 121 CCAGAGGACGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTCGCCGCGCTCGAG 180
Db 389 ccagaggacgcccctaacgcgcgcctccctggcgccacccctgtgtcccccgtctcgag 448
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCATTCTCTCCAG 240
Db 449 ccccttaactgacttcggagccctcgagctacgctctgagctcagcacttctctccag 508
QY 241 CAACACCCGCCGCTTC 258
Db 509 caacacccggcgcccttc 526

RESULT 9

US-09-699-998-9797
; Sequence 9797, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9797
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-699-998-9797

Query Match 100.0%; Score 258; DB 23; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCCAGTCCAGGAAACCGAACCCAGCAAAAGGCTTCTCTTCTGCTG 60
Db 269 gtccggcgccagctgccagtcgaggaacccagcaaaagcttcttcttctgctg 328
QY 61 CTCACATCGTCTTCTGCGACATCTGATGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120
Db 329 ctccacatcgcttctgcccagatctctgctgaagagggtgtgccggcgccctgcct 388
QY 121 CCAGAGGACGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTCGCCGCGCTCGAG 180
Db 389 ccagaggacgcccctaacgcgcgcctccctggcgccacccctgtgtcccccgtctcgag 448
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCATTCTCTCCAG 240
Db 449 ccccttaactgacttcggagccctcgagctacgctctgagctcagcacttctctccag 508
QY 241 CAACACCCGCCGCTTC 258
Db 509 caacacccggcgcccttc 526

RESULT 10
US-09-710-286-3772
; Sequence 3772, Application US/09710286
; GENERAL INFORMATION:
; APPLICANT: Gearling, David P.
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2005-001
; CURRENT APPLICATION NUMBER: US/09/710,286
; CURRENT FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: 60/164,255
; PRIOR FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 4115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3772
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-710-286-3772

Query Match 100.0%; Score 258; DB 24; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGCGCCAGCTGCCAGTCGAGGAAACCCAGCCAAAGAGCTTCCTCTCTGCTG 60
DB 269 gtccggcgccagctgccagtcgaggaacccagccaaagagctctctcttctgctg 328
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGCGGCGCCCTCGGACCTTCGACACTTCC 120
DB 329 ctccaccatcgctctctgccagatccctgagccctcgagcccccctgctccctc 388
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGCGCGCCACCCCTGTCGCGCGCCCTCG 180
DB 389 ccagagagcgcccttaacgcgcgcctcctcgagccctcgagcccccctgctcc 448
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACCTTCGACCTTCGACACTTCCCTCC 240
DB 449 ccccttaatctgacttcggagccctcgagccctcgagcccccctgctccctc 508
QY 241 CAACACCGCGCGCCTTC 258
DB 509 caacacccggcgcccttc 526

Query Match 100.0%; Score 258; DB 24; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGCGCCAGCTGCCAGTCGAGGAAACCCAGCCAAAGAGCTTCCTCTCTGCTG 60
DB 269 gtccggcgccagctgccagtcgaggaacccagccaaagagctctctcttctgctg 328
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGCGGCGCCCTCGGACCTTCGACACTTCC 120
DB 329 ctccaccatcgctctctgccagatccctgagccctcgagcccccctgctccctc 388
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGCGCGCCACCCCTGTCGCGCGCCCTCG 180
DB 389 ccagagagcgcccttaacgcgcgcctcctcgagccctcgagcccccctgctcc 448
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACCTTCGACCTTCGACACTTCCCTCC 240
DB 449 ccccttaatctgacttcggagccctcgagccctcgagcccccctgctccctc 508
QY 241 CAACACCGCGCGCCTTC 258
DB 509 caacacccggcgcccttc 526

Query Match 100.0%; Score 258; DB 24; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGCGCCAGCTGCCAGTCGAGGAAACCCAGCCAAAGAGCTTCCTCTCTGCTG 60
DB 269 gtccggcgccagctgccagtcgaggaacccagccaaagagctctctcttctgctg 328
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGCGGCGCCCTCGGACCTTCGACACTTCC 120
DB 329 ctccaccatcgctctctgccagatccctgagccctcgagcccccctgctccctc 388
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGCGCGCCACCCCTGTCGCGCGCCCTCG 180
DB 389 ccagagagcgcccttaacgcgcgcctcctcgagccctcgagcccccctgctcc 448
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACCTTCGACCTTCGACACTTCCCTCC 240
DB 449 ccccttaatctgacttcggagccctcgagccctcgagcccccctgctccctc 508
QY 241 CAACACCGCGCGCCTTC 258
DB 509 caacacccggcgcccttc 526

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGCGCCAGCTGCCAGTCGAGGAAACCCAGCCAAAGAGCTTCCTCTCTGCTG 60
DB 269 gtccggcgccagctgccagtcgaggaacccagccaaagagctctcttcttctgctg 328
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGCGGCGCCCTCGGACCTTCGACACTTCC 120
DB 329 ctccaccatcgctctctgccagatccctgagccctcgagcccccctgctccctc 388
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGCGCGCCACCCCTGTCGCGCGCCCTCG 180
DB 389 ccagagagcgcccttaacgcgcgcctcctcgagccctcgagcccccctgctcc 448
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACCTTCGACCTTCGACACTTCCCTCC 240
DB 449 ccccttaatctgacttcggagccctcgagccctcgagcccccctgctccctc 508
QY 241 CAACACCGCGCGCCTTC 258
DB 509 caacacccggcgcccttc 526

RESULT 12
US-09-721-588-4548
; Sequence 4548, Application US/09721588
; GENERAL INFORMATION:
; APPLICANT: Gearling, David P.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Villevial, Jean-Luc
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2046-001
; CURRENT APPLICATION NUMBER: US/09/721,588
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: 60/167,381
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 5410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4548
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-721-588-4548

Query Match 100.0%; Score 258; DB 24; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGCGCCAGCTGCCAGTCGAGGAAACCCAGCCAAAGAGCTTCCTCTCTGCTG 60
DB 269 gtccggcgccagctgccagtcgaggaacccagccaaagagctctcttcttctgctg 328
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGCGGCGCCCTCGGACCTTCGACACTTCC 120
DB 329 ctccaccatcgctctctgccagatccctgagccctcgagcccccctgctccctc 388
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGCGCGCCACCCCTGTCGCGCGCCCTCG 180
DB 389 ccagagagcgcccttaacgcgcgcctcctcgagccctcgagcccccctgctcc 448
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACCTTCGACCTTCGACACTTCCCTCC 240
DB 449 ccccttaatctgacttcggagccctcgagccctcgagcccccctgctccctc 508
QY 241 CAACACCGCGCGCCTTC 258
DB 509 caacacccggcgcccttc 526

RESULT 13
US-09-726-787-2982

```
; Sequence 2982, Application US/09726787
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2010-001
; CURRENT APPLICATION NUMBER: US/09/726,787
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/168,132
; PRIOR FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 3241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2982
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-787-2982
```

```
Query Match 100.0%; Score 258; DB 55; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GTCCGGCCAGCTGCCAGTCCAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 60
DB 269 gtccggccagctgccagtcagagaaacccagccaaagagcttcttctgctg 328

QY 61 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCGCTGCT 120
DB 329 ctccacatctgttctgccagatcctgaggtgagaggggtgtgccggccctgct 388

QY 121 CCAGAGGACGCCCTAACCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 180
DB 389 ccagaggacgcccctaaacgcgcctccctggcgccacccctggtgcccgctcgag 448

QY 181 CCCTTTAATCTGACTTCGGACCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCC 240
DB 449 ccctttaatctgacttcggacccctcggaactacgctctggacctcagcacttctccag 508

QY 241 CAACACCGCGCGCTTC 258
DB 509 caacaccggcgcccttc 526
```

```
RESULT 14
US-09-699-998-1778
; Sequence 1778, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1778
; LENGTH: 706
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(706)
; OTHER INFORMATION: n = A,T,C or G
US-09-699-998-1778
```

```
Query Match 100.0%; Score 258; DB 23; Length 706;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
```

```
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCGGCCAGCTGCCAGTCCAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 60
DB 195 gtccggccagctgccagtcagagaaacccagccaaagagcttcttctgctg 254

QY 61 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCGCTGCT 120
DB 255 ctccacatctgttctgccagatcctgaggtgagaggggtgtgccggccctgct 314

QY 121 CCAGAGGACGCCCTAACCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 180
DB 315 ccagaggacgcccctaaacgcgcctccctggcgccacccctgtgcccgctcgag 374

QY 181 CCCTTTAATCTGACTTCGGACCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCC 240
DB 375 ccctttaatctgacttcggacccctcggaactacgctctggacctcagcacttctccag 434

QY 241 CAACACCGCGCGCTTC 258
DB 435 caacaccggcgcccttc 452
```

```
RESULT 15
US-09-699-998-10469
; Sequence 10469, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10469
; LENGTH: 803
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-699-998-10469
```

```
Query Match 100.0%; Score 258; DB 23; Length 803;
Best Local Similarity 100.0%; Pred. No. 7.8e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCGGCCAGCTGCCAGTCCAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 60
DB 184 gtccggccagctgccagtcagagaaacccagccaaagagcttcttctgctg 243

QY 61 CTCACCATCTGCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCGCTGCT 120
DB 244 ctccacatctgttctgccagatcctgaggtgagaggggtgtgccggccctgct 303

QY 121 CCAGAGGACGCCCTAACCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 180
DB 304 ccagaggacgcccctaaacgcgcctccctggcgccacccctgtgtcccgctcgag 363

QY 181 CCCTTTAATCTGACTTCGGACCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCC 240
DB 364 ccctttaatctgacttcggacccctcggaactacgctctggacctcagcacttctccag 423

QY 241 CAACACCGCGCGCTTC 258
DB 424 caacaccggcgcccttc 441
```

```
Search completed: January 30, 2001, 22:04:50
Job time: 13154 sec
```



```

; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLHR, HOUBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1120 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 80..913
; US-08-646-715-8

```

```

Query Match      13.0%; Score 33.6; DB 1; Length, 1120;
Best Local Similarity 51.0%; Pred. No. 1.2;
Matches 104; Conservative 0; Mismatches 99; Indels 1; Gaps 1;

QY 34 CCAGCCAAAGGCTTCTTTCTGCTGCCTCACCACATCTCTTCTGCCAGATCTCTGATGGCT 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 866 CCACCTCAACGCCAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 94 GAAGAGGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 807 GAACCGGATGCGCCAGGATGCTGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 154 CCACCCCTGTGTCCTGCTGCGAGCCCTTTAATCTGACATTCGAGGCGCTTCGGACTAC 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 747 AGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 214 GCTCTGGACCTCAGCACTTTCCTC 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 687 CCGCCGAGCTGCCACCGTTTTC 564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-07-705-490-1/c
; Sequence 1, Application US/07705490
; Patent No. 6107025
; GENERAL INFORMATION:
; APPLICANT: Caskey, C. T.
; APPLICANT: Nelson, David L.
; APPLICANT: Pieretti, Maura

```

```

; APPLICANT: Warren, Stephen T.
; APPLICANT: Oosira, Ben A.
; TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas D. Paul
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/705,490
; FILING DATE: 19910708
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3765 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-07-705-490-1

```

```

Query Match      12.4%; Score 32; DB 3; Length 3765;
Best Local Similarity 49.5%; Pred. No. 4.1;
Matches 109; Conservative 0; Mismatches 110; Indels 1; Gaps 1;

QY 28 CCAGACCCAGCCAAAGGCTTCTTTCTGCTGCTCACCATCTCTTCTGCCAGATCTCTG 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 222 CCACCTTCCACCACGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 88 ATGGCTGAAGAGGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 162 CCCCCGAGAGTGGGCTGCGCGCGCTCGAGGC-CCAGCCGCGCGCGCGCGCGCGCGCG 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 148 CTGGGCGCCACCCCTGTGTCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 103 GCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 208 GACTAGGCTCTGAGCCTCAGCACTTCTCTCCAGCAACACC 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 43 GCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
US-08-455-073A-1/c
; Sequence 1, Application US/08455073A
; Patent No. 5876949
; GENERAL INFORMATION:
; APPLICANT: Gideon Dreyfuss
; APPLICANT: Mikiko C. Siomi
; APPLICANT: Yan Zhang
; TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods
; TITLE OF INVENTION: Of Making And Using The Same
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:

```

RESULT 10

US-08-859-201-1/c
 ; Sequence 1, Application US/08859201
 ; Patent No. 5880261
 ; GENERAL INFORMATION:
 ; APPLICANT: Waeber, G
 ; APPLICANT: Nicod, P
 ; TITLE OF INVENTION: Transcription Factor Islet-Brain 1 (IB1)
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Luann Cserf
 ; STREET: 750 Arimo Avenue
 ; CITY: Oakland
 ; STATE: California
 ; COUNTRY: USA

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/859,201
 FILING DATE: 20-MAY-1997

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9706731.8
 FILING DATE: 03-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9709920.4
 FILING DATE: 15-MAY-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Luann Cserf
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: ME A9702

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2953 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:

NAME/KEY: CDS
 LOCATION: 108..2252

US-08-859-201-1

Query Match 13.5%; Score 34.8; DB 2; Length 2953;
 Best Local Similarity 62.8%; Pred. No. 0.69; Mismatches 32; Indels 0; Gaps 0;
 Matches 54; Conservative 0;
 QY 98 AGCGTGTCCGGGCGCCCTCCAGAGAGCCCTTAACGCCGATCCCTGGGCCCA 157
 DB 2539 AGGATGTGACGGCGCTCTGCTAGGGACCCCTTTGTGGCCAGATCCCTGACCCCA 2480
 QY 158 CCCTGTGTCCCGCTCCCTGAGCCC 183
 DB 2479 CCCCAGCCCTGCCCTTCCCGAGCCCC 2454

RESULT 11

US-08-188-582-8/c
 ; Sequence 8, Application US/08188582
 ; Patent No. 5534410
 ; GENERAL INFORMATION:
 ; APPLICANT: Tjian, Robert
 ; APPLICANT: Comai, Lucio
 ; APPLICANT: Dynlacht, Brian D.
 ; APPLICANT: Huey, Timothy
 ; APPLICANT: Siegfried

; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 ; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 36
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/188,582
 FILING DATE: 28-JAN-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1120 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 80...913
 US-08-188-582-8

Query Match 13.0%; Score 33.6; DB 1; Length 1120;
 Best Local Similarity 51.0%; Pred. No. 1.2;
 Matches 104; Conservative 0; Mismatches 99; Indels 1; Gaps 1;
 QY 34 CCAGCCAAAGGCTTCTTTCTGCTGCATCCATCTTCTGCGAGATCCTGATGGCT 93
 DB 866 CCATCCCAACGCCAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 808
 QY 94 GAAGAGGTGTCCGGGCGCCCTGCTCCAGAGAGCGCCCTTAACGCCGATCCCTGGCG 153
 DB 807 GAACCGATGCCCGAGCGATGCTGCCACGCCCGCGATCGCTGCCACCTCCATCTTG 748
 QY 154 CCCACCCCTGTGTCCCGCTCTCTCGAGCCCTTTAATCTGACITTCGAGCCCTCGGACTAC 213
 DB 747 AGATCTCGCGCGCGGTGCTCTGCTCTTAACCTCTGACACCCCGCCCGCCAGAT 688
 QY 214 GCTCTGGACCTCAGCATTTCTCTC 237
 DB 687 CCGCGGAGCTGCCACCGCTTTTC 664

RESULT 12

US-08-646-715-8/c
 ; Sequence 8, Application US/08646715
 ; Patent No. 5637686
 ; GENERAL INFORMATION:
 ; APPLICANT: Tjian, Robert
 ; APPLICANT: Comai, Lucio
 ; APPLICANT: Dynlacht, Brian D.
 ; APPLICANT: Huey, Timothy
 ; APPLICANT: Ruppert, Siegfried
 ; APPLICANT: Tanese, Naoko
 ; APPLICANT: Wang, Edith

US-08-826-246-9

Query Match 100.0%; Score 258; DB 3; Length 1228;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCGGCGCAGTGCATTCGAGGAGCAACCCAGCCAAAGGCTTCTTTCTGCTG 60
Db 211 GTCCGGCGCAGTGCATTCGAGGAGCAACCCAGCCAAAGGCTTCTTTCTGCTG 270
OY 61 CTCACCATCGTCTTCTGCGAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCCCTGCT 120
Db 271 CTCACCATCGTCTTCTGCGAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCCCTGCT 330
OY 121 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCGCGCCCTGCT 180
Db 331 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCGCGCCCTGCT 390
OY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCTCTCCAG 240
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCTCTCCAG 450
OY 241 CAACACCGCGCGCCTTC 258
Db 451 CAACACCGCGCGCCTTC 468

RESULT 2

US-08-944-495-9
; Sequence 9, Application US/08944495
; Patent No. 6087477

GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELETYPE: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1228 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:

NAME/KEY: Coding Sequence
; LOCATION: 1...468
; OTHER INFORMATION:
US-08-944-495-9

Query Match 100.0%; Score 258; DB 3; Length 1228;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCGGCGCAGTGCATTCGAGGAGCAACCCAGCCAAAGGCTTCTTTCTGCTG 60
Db 211 GTCCGGCGCAGTGCATTCGAGGAGCAACCCAGCCAAAGGCTTCTTTCTGCTG 270
OY 61 CTCACCATCGTCTTCTGCGAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCCCTGCT 120
Db 271 CTCACCATCGTCTTCTGCGAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCCCTGCT 330
OY 121 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCGCGCCCTGCT 180
Db 331 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCGCGCCCTGCT 390
OY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCTCTCCAG 240
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCTCTCCAG 450
OY 241 CAACACCGCGCGCCTTC 258
Db 451 CAACACCGCGCGCCTTC 468

RESULT 3

US-09-126-640-5
; Sequence 5, Application US/09126640A
; Patent No. 6099823
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/126,640A
; CURRENT FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 08/870,414
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 08/799,910
; EARLIER FILING DATE: 1997-02-13
; EARLIER APPLICATION NUMBER: 60/011,787
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1228
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-126-640-5

Query Match 100.0%; Score 258; DB 3; Length 1228;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCGGCGCAGTGCATTCGAGGAGCAACCCAGCCAAAGGCTTCTTTCTGCTG 60
Db 211 GTCCGGCGCAGTGCATTCGAGGAGCAACCCAGCCAAAGGCTTCTTTCTGCTG 270
OY 61 CTCACCATCGTCTTCTGCGAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCCCTGCT 120
Db 271 CTCACCATCGTCTTCTGCGAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCCCTGCT 330
OY 121 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCGCGCCCTGCT 180
Db 331 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCGCGCCCTGCT 390

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 21:28:30 ; Search time 49.1 Seconds
(without alignments)
846.830 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468
Perfect score: 258
Sequence: 1 GTCGGCGCCACTGCCAGT.....ACCAACACCCGCCGCCCTTC 258

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	100.0	1228	3	US-08-826-246-9
2	258	100.0	1228	3	US-08-944-495-9
3	258	100.0	1228	3	US-09-126-640-5
4	43	16.7	7218	1	US-08-232-463-14
5	39.4	15.3	2847	1	US-08-087-007-2
6	39.4	15.3	2847	3	US-08-483-433-2
7	39.4	15.3	2847	4	PCT-US92-05920-2
8	38.6	15.0	6354	3	US-09-058-389A-5
9	37.6	14.6	170	3	US-09-058-389A-13
10	34.8	13.5	2953	2	US-08-859-201-1
11	33.6	13.0	1120	1	US-08-188-582-8
12	33.6	13.0	1120	1	US-08-646-715-8
13	32	12.4	3765	3	US-07-705-490-1
14	32	12.4	4362	2	US-08-455-073A-1
15	31.8	12.3	599	1	US-08-584-226-3
16	31.6	12.2	2074	2	US-09-018-576-2
17	31.6	12.2	2074	3	US-09-018-576-11
18	31.6	12.2	2074	3	US-09-248-137-2
19	31.6	12.2	2074	3	US-09-248-137-11
20	31.4	12.2	1335	4	PCT-US91-06532-1
21	31.4	12.2	38506	3	US-09-320-878-19
22	30.8	11.9	797	1	US-08-332-467B-2
23	30.8	11.9	797	1	US-08-681-811-2
24	30.8	11.9	797	4	PCT-US93-12507-2
25	30.8	11.9	2680	3	US-09-063-035-1
26	30.6	11.9	7791	3	US-08-949-386-23
27	30.6	11.9	7791	3	US-08-450-562-23
28	30.6	11.9	7808	3	US-08-949-386-22

c	29	30.6	11.9	7808	3	US-08-450-562-22
	30	30.2	11.7	22481	4	PCT-US95-07201-43
	31	30	11.6	3688	5	5248670-4
	32	29.8	11.6	536	1	US-08-329-704-1
	33	29.8	11.6	536	2	US-08-472-604-1
	34	29.8	11.6	536	2	US-08-486-117-1
	35	29.8	11.6	536	2	US-08-865-337A-3
	36	29.8	11.6	2196	2	US-08-865-337A-2
	37	29.8	11.6	8299	1	US-08-462-014-2
	38	29.8	11.6	8299	3	US-08-923-137-3
	39	29.6	11.5	693	1	US-08-168-091A-3
c	40	29.6	11.5	1219	4	PCT-US93-06251-11
	41	29.6	11.5	1313	1	US-08-176-427B-7
	42	29.6	11.5	1313	2	US-08-356-060A-4
	43	29.6	11.5	1313	3	US-08-460-900C-4
	44	29.6	11.5	1489	3	US-08-836-582-1
	45	29.6	11.5	2485	1	US-08-424-424B-1

ALIGNMENTS

RESULT 1
US-08-826-246-9
; Sequence 9, Application US/08826246
; Patent No. 6048709
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,246
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-078-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1228 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...468
; OTHER INFORMATION:

```
unclassified
REFERENCE 1 (bases 1 to 297)
AUTHORS  Sibson,D.R. and Hadfield,K.M.
TITLE     HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,
          PLACENTA OR BONE MARROW AND THEIR USE
JOURNAL   Patent: EP 0587279-A 16-MAR-1994;
          MEDICAL RES COUNCIL (GB)
FEATURES  Location/Qualifiers
            source
              1..297
                /organism="unidentified"
                /db_xref="taxon:32644"
BASE COUNT 66 a 117 c 61 g 52 t 1 others
ORIGIN

Query Match      17.1%; Score 44; DB 81; Length 297;
Best Local Similarity 77.9%; Pred. NO. 4.3;
Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY  2  TCCGGCCCGCAGTGCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTGC 61
      |||||
DB  225 TCCGGCCCGCAGTGCAGTTAAGGACCAACCTCGCCAGAAGGCTTCTTTTCGGGTTGA 284
      |||||

OY  62  TCACCATC 69
      |||||
DB  285 TCGCCATC 292
```

Search completed: January 30, 2001, 19:24:37

Job time: 25963 sec


```
* 124276 124375: gap of unknown length
* 124376 126296: contig of 1921 bp in length
* 126297 126396: gap of unknown length
* 126397 129180: contig of 2784 bp in length
* 129181 129280: gap of unknown length
* 129281 132785: contig of 3405 bp in length
* 132686 132785: gap of unknown length
* 132786 134635: contig of 1850 bp in length
* 134636 134735: gap of unknown length
* 134736 137197: contig of 2462 bp in length
* 137198 137297: gap of unknown length
* 137298 139569: contig of 2272 bp in length
* 139570 139669: gap of unknown length
* 139670 142014: contig of 2345 bp in length
* 142015 142114: gap of unknown length
* 142115 144074: contig of 1960 bp in length
* 144075 144174: gap of unknown length
* 144175 146482: contig of 2308 bp in length
* 146483 146582: gap of unknown length
* 146583 148126: contig of 1544 bp in length
* 148127 148226: gap of unknown length
* 148227 150342: contig of 2116 bp in length
* 150343 150442: gap of unknown length
* 150443 151791: contig of 1349 bp in length
* 151792 151891: gap of unknown length
* 151892 153769: contig of 1878 bp in length
* 153770 153869: gap of unknown length
* 153870 156824: contig of 2955 bp in length
* 156825 156924: gap of unknown length
* 156925 158316: contig of 1392 bp in length
* 158317 158416: gap of unknown length
* 158417 160019: contig of 1603 bp in length
* 160020 160119: gap of unknown length
* 160120 161830: contig of 1711 bp in length
* 161831 161930: gap of unknown length
* 161931 163680: contig of 1750 bp in length
* 163681 163780: gap of unknown length
* 163781 165772: contig of 1992 bp in length
* 165773 165872: gap of unknown length
* 165873 167356: contig of 1484 bp in length
* 167357 167456: gap of unknown length
* 167457 169605: contig of 2149 bp in length
* 169606 169705: gap of unknown length
* 169706 171781: contig of 2076 bp in length
* 171782 171881: gap of unknown length
* 171882 173350: contig of 1469 bp in length
* 173351 173450: gap of unknown length
* 173451 175539: contig of 2089 bp in length
* 175540 175639: gap of unknown length
* 175640 176975: contig of 1336 bp in length
* 176976 177075: gap of unknown length
* 177076 178320: contig of 1445 bp in length
* 178321 178620: gap of unknown length
* 178621 179634: contig of 1014 bp in length
* 179635 179734: gap of unknown length
* 179735 181530: contig of 1796 bp in length
* 181531 181630: gap of unknown length
* 181631 182891: contig of 1261 bp in length
* 182892 182991: gap of unknown length
* 182992 184207: contig of 1216 bp in length
* 184208 184307: gap of unknown length
* 184308 185488: contig of 1181 bp in length
* 185489 185588: gap of unknown length
* 185589 186739: contig of 1151 bp in length
* 186740 186839: gap of unknown length
* 186840 187868: contig of 1029 bp in length.
FEATURES          Location/Qualifiers
Source            1..187868
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
                  /clone="RP23-329K24"
BASE COUNT       46889 a 45196 c 44010 g 45923 t 5850 others
ORIGIN
```

```
Query Match          54.0%; Score 139.2; DB 62; Length 187868;
Best Local Similarity 74.8%; Pred. No. 2e-21;
Matches 193; Conservative 0; Mismatches 53; Indels 12; Gaps 1;

QY 13 CTGCCAGTCGAGAACCCGACGCAAAAGGCTTCTCTTTCTGCTGCTCACCATCGTC 72
Db 39433 CTACCAACCGAGGACCCCAACATTGCCAAGAGGTCCTCTTTCTCTGTTCCGATCATC 39492
QY 73 TTTCGCAGATCCTGATGCTGAAGAGGTTGTCCCGCGCCCTGCTCCAGAGGACGCC 132
Db 39493 TTTCGCAGATTTGATGCTGAAGAGGTTGTCTCCAGCCCTGCTCCGAGGATGCT 39552
QY 133 CCTAACGCCGCATCCC-----TGGCGCCACACCCCTGTGTGCCCGCTCTCGAG 180
Db 39553 ACCAGCGCGTGACACCTGAGCCCATTTCTGCGCCCATTTACTGCGCCCTCGGTCGAG 39612
QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTTCCTCCAG 240
Db 39613 CCTTTGAACTGACCTCGGAGTCTCGGACTATGCGCTGGATCTTAAGCTTTCTCCAG 39672
QY 241 CAACACCGCGCGCCTTC 258
Db 39673 CAACATCCGCGCGCCTTC 39690

RESULT 14
A75448
LOCUS       A75448      297 bp      DNA
DEFINITION Sequence 1134 from Patent WO9401548.
ACCESSION  A75448
VERSION    A75448.1 GI:6065525
KEYWORDS   unidentified.
SOURCE     unidentified.
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 297)
AUTHORS   Sibson,D.R. and Gross,J.
TITLE     HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE,
          PLACENTA OR BONE MARROW
JOURNAL   Patent: WO 9401548-A 20-JAN-1994;
          MEDICAL RES COUNCIL (GB); SIBSON DAVID ROSS (GB)
FEATURES   source
            1..297
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT 66 a 117 c 61 g 52 t 1 others
ORIGIN

Query Match          17.1%; Score 44; DB 81; Length 297;
Best Local Similarity 77.9%; Pred. No. 4.3;
Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 TCCGGCGCCAGCTGCCAGTCGAGGAACCCGACCCCAACCCCAAGGCTTCTCTTCTCGTCC 61
Db 225 TCCGGCGCAGTTGCCAGTTAAGGACCAACCTCGCCAGAGGCTTCTTTTCGGGTGA 284
QY 62 TCACCATC 69
Db 285 TCGCCATC 292

RESULT 15
A78427
LOCUS       A78427      297 bp      DNA
DEFINITION Sequence 1134 from Patent EP0587279.
ACCESSION  A78427
VERSION    A78427.1 GI:6090092
KEYWORDS   unidentified.
SOURCE     unidentified.
ORGANISM   unidentified
```


COMMENT

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 0
Center clone name: RPCI-23_128D3

Summary Statistics
Consensus quality: 139878 bases at least Q40
Consensus quality: 165603 bases at least Q30
Consensus quality: 177139 bases at least Q20
Estimated insert size: 183300; agarose-fp estimation
Estimated insert size: 200576; sum-of-contigs estimation
Quality coverage: 3.48 in Q20 bases; agarose-fp estimation
Quality coverage: 3.18 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1175: contig of 1175 bp in length
1176 1275: gap of unknown length
1276 2563: contig of 1288 bp in length
2564 2663: gap of unknown length
2664 3906: contig of 1243 bp in length
3907 4006: gap of unknown length
4007 5224: contig of 1218 bp in length
5225 5324: gap of unknown length
5325 6636: contig of 1312 bp in length
6637 9196: contig of 2460 bp in length
9197 9296: gap of unknown length
9297 10723: contig of 1427 bp in length
10724 10823: gap of unknown length
10824 12852: contig of 2029 bp in length
12853 12953: gap of unknown length
12953 16025: contig of 3073 bp in length
16026 16125: gap of unknown length
16126 19096: contig of 2971 bp in length
19097 19196: gap of unknown length
19197 21750: contig of 2554 bp in length
21751 21850: gap of unknown length
21851 26579: contig of 4729 bp in length
26580 26679: gap of unknown length
26680 32472: contig of 5793 bp in length
32473 32572: gap of unknown length
32573 36994: contig of 4422 bp in length
36995 37094: gap of unknown length
37095 44071: contig of 6977 bp in length
44072 44171: gap of unknown length
44172 50727: contig of 6556 bp in length
50728 50827: gap of unknown length
50828 57204: contig of 6377 bp in length
57205 57304: gap of unknown length
57305 65076: contig of 7772 bp in length
65077 65177: gap of unknown length
65178 71824: contig of 6648 bp in length
71825 71924: gap of unknown length
71925 78033: contig of 6111 bp in length
78036 78135: gap of unknown length
78136 87838: contig of 9703 bp in length
87839 87939: gap of unknown length
87940 94924: contig of 6986 bp in length
94925 95024: gap of unknown length
95025 101188: contig of 6164 bp in length
101189 101288: gap of unknown length
101289 108925: contig of 7637 bp in length
108926 109025: gap of unknown length

* 109026 119738: contig of 10713 bp in length
* 119739 119838: gap of unknown length
* 119839 129876: contig of 10038 bp in length
* 129877 129976: gap of unknown length
* 129977 143243: contig of 13267 bp in length
* 143244 161164: gap of unknown length
* 161165 161264: contig of 17821 bp in length
* 161265 180316: gap of unknown length
* 180317 180416: contig of 19052 bp in length
* 180417 203476: contig of 23060 bp in length.
FEATURES
Location/Qualifiers
1..203476
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-128D3"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 51748 a 48714 c 48138 g 51580 t 3296 others
ORIGIN
Query Match 58.6%; Score 151.2; DB 70; Length 203476;
Best Local Similarity 75.9%; Pred. No. 4.7e-24;
Matches 205; Conservative 0; Mismatches 53; Indels 12; Gaps 1;
QY 1 GTCGGGGCCAGCTGCCAGTCGAGGAACCCAGCAACCCAGCAAAAGGCTTCTTTCTGCTG 60
|||||
DB 92875 GTCGGGGCCAGCTACCAACCGGAGGAGCAACCAATGCAAGAGGCTCTCTTCTCCTG 92934
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGTGAAGAGGCTGTCGGCGGCCCTGCT 120
|||||
DB 92935 TTCGCCATCATCTTCTGCCAGATTTTGATGGCTGAAGAGGCTGTCGACGCCCTGGCT 92994
QY 121 CCAGAGGAGCGCCCTAACCGCGCATCCC-----TGGCGCCCAACCCCTGTGTC 168
|||||
DB 92995 CCGGAGGATGCTACCAAGCGCGTGACACCTGAGCCCATTTCTGCGCCCATTTACTCGCGCC 93054
QY 169 CCGTCTCTGAGCGCTTTAATCTGACTTCGGAGCGCTCGGAGCGCTCTGACCTCAGC 228
|||||
DB 93055 CCGTCTCTGAGCGCTTTGAACCTGACCTCGGAGTCTCGAGCTATCGCTGCTGTTAA 93114
QY 229 ACTTCTCTCCAGCAACCCCGCGCTTC 258
|||||
DB 93115 GCTTCTCCAGCAACATCCCGCGCTTC 93144
RESULT 12
MMGLY96 1938 bp mRNA ROD 11-MAR-1993
LOCUS M.musculus gly96 mRNA.
DEFINITION X67644
ACCESSION X67644.1 GI:287803
VERSION gly96 gene; glycosylated protein.
KEYWORDS house mouse.
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1938)
AUTHORS Lau, L.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1992) L. Lau, University of Illinois College of
Med., Dept. of Genetics, M/C 669, 808 South Wood Street, Chicago,
IL 60612, USA
REFERENCE
2 (bases 1 to 1938)
AUTHORS Charles, C.H., Yoon, J.K., Simske, J.S. and Lau, L.F.
TITLE Genomic structure; cDNA sequence, and expression of gly96, a growth
factor-inducible immediate-early gene encoding a short-lived
glycosylated protein
JOURNAL Oncogene 8 (3), 797-801 (1993)
MEDLINE 93173526
FEATURES Location/Qualifiers
1..1938
source /organism="Mus musculus"

Db 61000 GTCCGGCGCCAGGTGCAGTTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 61059
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTTCCT 120
Db 61060 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTTCCT 61119
Qy 121 CCAGAGACGCCCTTAACGGCGGATCCTGTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG 180
Db 61120 CCAGAGACGCCCTTAACGGCGGATCCTGTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG 61179
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCGCCAG 240
Db 61180 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCGCCAG 61239
Qy 241 CAACACCCGCCCTTC 258
Db 61240 CAACACCCGCCCTTC 61257
RESULT 8
AP000512 DNA 30-MAR-2000
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
DEFINITION section 11/20.
ACCESSION AP000512
VERSION AP000512.1 GI:5926699
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (sites)
Shiina,S., Tamliya,G., Oka,A. and Inoko,H.
Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region
Published Only in DataBase (1999) In press
2 (bases 1 to 200000)
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
Direct Submission
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www.alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is conducted by Tokai University as a JST sequencing
Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive Web site (<http://www.alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www.alis.tokyo.jst.go.jp.
FEATURES
source
1..200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
101440..101648
/note="stSG4264:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4566033"
complement(114197..114328)
/standard_name="D6S2113"
/note="SHGC-13451:The location is between each flanking
site of PCR primers."
/db_xref="GDB:738998"
complement(114300..114430)
/note="WI-18721:The location is between each flanking site

of PCR primers."
/db_xref="GDB:4574312"
117236..117362
/note="SGC34396:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4580073"
complement(117712..117836)
/standard_name="D6S2041"
/note="SHGC-10364:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735115"
complement(117961..118093)
/note="SGC32809:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4584619"
142184..142339
/note="KIAA0170:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4586652"
complement(142215..142343)
/note="WI-15445:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4581086"
complement(153977..154114)
/note="SGC34551:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4582352"
complement(165191..165295)
/standard_name="D6S1376"
/note="downstream/upstream:The location is between each
flanking site of PCR primers."
/db_xref="GDB:580497"
188925..188989
/note="pml216:The location is between each flanking site
of PCR primers."
/db_xref="GDB:594325"
complement(197650..197804)
/note="Cda01e04:The location is between each flanking site
of PCR primers."
/db_xref="GDB:443319"
BASE COUNT 52458 a 48336 c 49153 g 50053 t
ORIGIN
Query Match 100.0%; Score 258; DB 51; Length 200000;
Best Local Similarity 100.0%; Pred. No. 2.4e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTCCGGCGCCAGGTGCAGTTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
Db 97985 GTCCGGCGCCAGGTGCAGTTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 98044
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTTCCT 120
Db 98045 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTTCCT 98104
Qy 121 CCAGAGACGCCCTTAACGGCGGATCCTGTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG 180
Db 98105 CCAGAGACGCCCTTAACGGCGGATCCTGTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG 98164
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCGCCAG 240
Db 98165 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCGCCAG 98224
Qy 241 CAACACCCGCCCTTC 258
Db 98225 CAACACCCGCCCTTC 98242
RESULT 9
S81914 1223 bp mRNA PRI 12-AUG-1996
LOCUS S81914
DEFINITION IEX-1-radiation-Inducible immediate-early gene (human, placenta,

```
/rpt_family="L1"
13869. .14279
/rpt_family="Alu"
18033. .18284
/rpt_family="Alu"
19178. .19451
/rpt_family="Alu"
19458. .19749
/rpt_family="Alu"
20862. .21002
/rpt_family="MERS"
21586. .21832
/rpt_family="Alu"
22231. .22521
/rpt_family="Alu"
complement(22616. .22679)
/rpt_family="Alu"
complement(22869. .23145)
/rpt_family="Alu"
complement(23478. .23772)
/rpt_family="Alu"
24184. .24372
/rpt_family="Alu"
24529. .24813
/rpt_family="Alu"
complement(25060. .25342)
/rpt_family="Alu"
complement(25367. .25653)
/rpt_family="Alu"
25787. .26076
/rpt_family="Alu"
26491. .26762
/rpt_family="Alu"
26866. .27143
/rpt_family="Alu"
complement(27620. .27919)
/rpt_family="Alu"
29525. .29719
/rpt_family="MIR"
29762. .30333
/rpt_family="Alu"
30365. .30653
/rpt_family="Alu"
31430. .31722
/rpt_family="Alu"
31765. .31857
/rpt_family="Alu"
32265. .32563
/rpt_family="Alu"
complement(32710. .32858)
/rpt_family="Alu"
complement(33374. .33584)
/rpt_family="MER2"
complement(34547. .34595)
/rpt_family="MIR"
complement(36673. .36954)
/rpt_family="Alu"
complement(37002. .37230)
/rpt_family="Alu"
complement(40018. .40170)
/rpt_family="Alu"
40452
note="high quality variation versus 3' overlapping clone"
40509. .41109
/rpt_family="Alu"
42609. .42892
/rpt_family="Alu"
43240. .43370
/rpt_family="MIR"
43493. .43779
/rpt_family="Alu"
BASE COUNT 11505 a 10560 c 11226 g 10827 t
ORIGIN
```

```
Query Match          100.0%; Score 258; DB 9; Length 44118;
Best Local Similarity 100.0%; Pred. No. 3.3e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1  GTCCGGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGTG 60
      |||||||
Db 14923 GTCCGGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGTG 14982

Qy 61  CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGTCGCGCGCCCTGCTGCT 120
      |||||||
Db 14983 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGTCGCGCGCCCTGCTGCT 15042

Qy 121  CCAGAGGACGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTCTCCCGCTCTCCGAG 180
      |||||||
Db 15043 CCAGAGGACGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTCTCCCGCTCTCCGAG 15102

Qy 181  CCCTTTAATCTGACTTCGGAGCCCTTCGACTACGCTCTGGACCTCAGACATTTCTCCAG 240
      |||||||
Db 15103 CCCTTTAATCTGACTTCGGAGCCCTTCGACTACGCTCTGGACCTCAGACATTTCTCCAG 15162

Qy 241  CAACACCGCGCGCTTC 258
      |||||||
Db 15163 CAACACCGCGCGCTTC 15180

RESULT 7
AB023051 90244 bp DNA PRI 20-NOV-1999
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
DEFINITION clone:876L4, complete sequence.
ACCESSION AB023051
VERSION AB023051.1 GI:5672606
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:876L4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Shiina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Kuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Bahram,S. and Inoko,H.
TITLE Molecular dynamics of MHC genes unraveled by sequence analysis of
the 1,796,938-bp HLA class I region
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
MEDLINE 20027539
REFERENCE 2 (bases 1 to 90244)
AUTHORS Shiina,T. and Takishima,N.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
Shiina, Tokai University School of Medicine, Department of
Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail:tshiina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)
FEATURES
Location/Qualifiers
1..90244
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="876L4"
/map="6p21.3"

BASE COUNT 24562 a 21396 c 21572 g 22714 t
ORIGIN

Query Match          100.0%; Score 258; DB 8; Length 90244;
Best Local Similarity 100.0%; Pred. No. 2.8e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1  GTCCGGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGTG 60
```

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCAGCTGCAGTCGAGAACCCAGCCAAAGGCTTCCTTTCTGCTG 60
Db 914 GTCCGGCGCAGCTGCAGTCGAGAACCCAGCCAAAGGCTTCCTTTCTGCTG 973
QY 61 CTCACCATCCTCTCTCCAGATCCCTGATGGCTGAAGAGGCTGCCGGCGCCCTCCCT 120
Db 974 CTCACCATCCTCTCTCCAGATCCCTGATGGCTGAAGAGGCTGCCGGCGCCCTCCCT 1033
QY 121 CCAGAGAGCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGTGCCCGCTCCCTCGAG 180
Db 1034 CCAGAGAGCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGTGCCCGCTCCCTCGAG 1093
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCCGAG 240
Db 1094 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCCGAG 1153
QY 241 CAACACCGCGCCCTTC 258
Db 1154 CAACACCGCGCCCTTC 1171

RESULT 6

AC006165 44118 bp DNA PRI 08-DEC-1998
LOCUS Homo sapiens clone UWGC:y54c125 from 6p21, complete sequence.
DEFINITION AC006165
ACCESSION AC006165
VERSION AC006165.1 GI:3980464
KEYWORDS HTG.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 44118)
AUTHORS Janer, M., Guillaudoux, T., Vu, Q., Kutayavin, T., Harter, H. and Geraghty, D.E.

TITLE Large scale sequence analysis of the human MHC class I region
JOURNAL Unpublished (1998)
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024

REFERENCE 2 (bases 1 to 44118)
AUTHORS Geraghty, D.E. and Olson, M.V.

TITLE Direct Submission
JOURNAL Submitted (08-DEC-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145, Seattle, WA 98195

Contact: Daniel E. Geraghty (geraghty@fhcrc.org)

Overlapping Sequences:

5': UWGC:y54c283
3': UWGC:y54c222 (Genbank Accession: AC006049)

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 58.8%
DS or two chemistry coverage: 99.5%
Single stranded regions: 2

Sequence Validation:

This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below.

Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BgIII		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
6714.50	6452.00	1823.23	1794.00	5581.12	5401.00
1089.04	1085.00	1246.04	1228.00	10570.24	10123.00
1536.15	1528.00	5570.33	5372.00	10069.29	9626.00
3623.08	3537.00	1670.92	1630.00		
3784.08	3621.00	1270.76	1280.00		
6834.00	6689.00	4235.79	4109.00		
11234.36	10874.00	1171.00	1151.00		
762.43	762.00	3226.11	3096.00		
2727.38	2653.00	10501.64	10080.00		
1691.42	1684.00	4208.44	4052.00		
1295.60	1277.00				
746.87	734.00				
FEATURES					
Source					
1. 44118					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/chromosome="6"					
/map="6p21"					
/clone="CGM1:D100E8"					
/sub_clone="UWGC:y54c125"					
/cell_line="CGM1"					
/clone_lib="Wash U YAC library"					
218..405					
/rpt_family="MER3"					
400..999					
/rpt_family="Alu"					
1797..2378					
/rpt_family="Alu"					
2649..2956					
/rpt_family="Alu"					
complement(3372..3894)					
/rpt_family="Alu"					
3941..4070					
/rpt_family="Alu"					
complement(449..4724)					
/rpt_family="Alu"					
complement(5117..5389)					
/rpt_family="Alu"					
7668..7913					
/rpt_family="Alu"					
complement(7924..8226)					
/rpt_family="Alu"					
8250..8551					
/rpt_family="Alu"					
complement(8893..8973)					
/rpt_family="MIR"					
complement(11829..12096)					
/rpt_family="Alu"					
12416..12740					
/rpt_family="Alu"					
complement(13181..13247)					
/rpt_family="Alu"					
complement(13498..13647)					

RESULT 4
AF071596 1693 bp DNA PRI 10-NOV-1998
DEFINITION Homo sapiens apoptosis inhibitor (IEX-1L) gene, complete cds.
ACCESSION AF071596
VERSION AF071596.1 GI:3851531
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1693)
AUTHORS Wu, M.X., Ao, Z., Prasad, K.V., Wu, R. and Schlossman, S.F.
TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival
JOURNAL Science 281 (5379), 998-1001 (1998)
MEDLINE 98369175
REFERENCE 2 (bases 1 to 1693)
AUTHORS Wu, M.X. and Ao, Z.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1998) Tumor Immunology, Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
source
1..1693
/organism="Homo sapiens"
/db_xref="taxon:9606"
359..1693
/gene="IEX-1L"
/product="apoptosis inhibitor"
359..1693
/gene="IEX-1L"
385..966
/gene="IEX-1L"
/note="immediate early response protein"
/codon_start=1
/product="apoptosis inhibitor"
/protein_id="AAC72344.1"
/db_xref="GI:3851532"
/translation="MCHSRSCPTMTILOAPTAPSTIPGPRKSGPEITFDPLPEP
AAAPAGPSASGRHRRSRVLPVPRVLRQLPVEEPNPKRLI.FILLTIVFCOII.MAE
EGVPALPPEDAPNAASLAPTVPSPVLEPNTSEPSDYALDLSTFLQHPAAF"
BASE COUNT 348 a 529 c 456 g 360 t
ORIGIN

Query Match 100.0%; Score 258; DB 10; Length 1693;
Best Local Similarity 100.0%; Pred. No. 6.4e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCGGCCAGCTCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
|||||
DB 706 GTCCGGCCAGCTCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 765
|||||
OY 61 CTCACCATCGTCTTCTGCGAGATCCTGATGCTGAAGAGGCTGTGCCGCCGCCCTGCT 120
|||||
DB 766 CTCACCATCGTCTTCTGCGAGATCCTGATGCTGAAGAGGCTGTGCCGCCGCCCTGCT 825
|||||
OY 121 CCAGAGGAGCCCTTAAGCCGCCATCCCTGCGGCCCCACCCCTCTCTCCCGCTCTCCAG 180
|||||
DB 826 CCAGAGGAGCCCTTAAGCCGCCATCCCTGCGGCCCCACCCCTCTCTCCCGCTCTCCAG 885
|||||
OY 181 CCCTTAATCTGACTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCTCCAG 240
|||||
DB 886 CCCTTAATCTGACTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCTCCAG 945
|||||
OY 241 CAACACCGGCCCTTC 258
|||||
DB 946 CAACACCGGCCCTTC 963
|||||

RESULT 5

Query Match 100.0%; Score 258; DB 53; Length 1864;
Best Local Similarity 100.0%; Pred. No. 6.3e-47;

HSPRG1
LOCUS 1864 bp DNA PRI 06-NOV-1998
DEFINITION H.sapiens PRG1 gene.
ACCESSION X96438
VERSION X96438.1 GI:2440072
KEYWORDS PRG1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1864)
AUTHORS Trauzold, A.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1996) A. Trauzold, Laboratory of Molecular Gastroenterology, 1st Dept. of Medicine, University of Kiel, Schittenhelmstrasse 12, Kiel, D-24105, FRG
REMARK revised by [4]
REFERENCE 2 (bases 1 to 508)
AUTHORS Schaefer, H., Trauzold, A., Lettau, P., Kalthoff, H., Koelsch, U.R. and Schmidt, W.E.
TITLE cDNA cloning and sequencing of a novel human early response gene and characterization of its expression in pancreatic carcinoma cells
JOURNAL Gastroenterology In press
REFERENCE 3 (bases 1 to 1864)
AUTHORS Schaefer, H.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1997) H. Schaefer, Trauzold, Laboratory of Molecular Gastroenterology, 1st Dept. of Medicine, University of Kiel, Schittenhelmstrasse 12, Kiel, D-24105, FRG
REFERENCE 4 (bases 1 to 1864)
AUTHORS Schaefer, H., Diebel, J., Arlt, A., Trauzold, A. and Schmidt, W.E.
TITLE The promoter of human p22/PACAP response gene 1 (PRG1) contains functional binding sites for the p53 tumor suppressor and for NF-kappaB
JOURNAL FEBS Lett. 436 (2), 139-143 (1998)
MEDLINE 98452926
COMMENT On Sep 27, 1997 this sequence version replaced gi:1515291.
FEATURES
source
1..1864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="818-4"
1..562
/gene="PRG1"
1..1864
/gene="PRG1"
Join(563..801,914..1864)
/gene="PRG1"
563..801
/gene="PRG1"
/number=1
Join(592..801,914..1174)
/gene="PRG1"
/codon_start=1
/protein_id="CAA65304.1"
/db_xref="GI:2440073"
/db_xref="SWISS-PROT:P46695"
/translation="MCHSRSCPTMTILOAPTAPSTIPGPRKSGPEITFDPLPEP
AAAPAGPSASGRHRRSRVLPVPRVLRQLPVEEPNPKRLI.FILLTIVFCOII.MAE
EGVPALPPEDAPNAASLAPTVPSPVLEPNTSEPSDYALDLSTFLQHPAAF"
802..913
/gene="PRG1"
/number=1
914..1864
/gene="PRG1"
/number=2
BASE COUNT 395 a 571 c 519 g 379 t
ORIGIN


```
/db_xref="taxon:9606"
/cell_type="leukocytes"
<!.if
/gene="dif-2"
1..1230
/gene="dif-2"
12..482
/gene="dif-2"
/codon_start=1
/product="Dif-2 protein"
/protein_id="CAA74886.1"
/db_xref="GI:2660551"
/db_xref="SWISS-PROT:P46695"
/translation="MCHSRCHPTMTILOAPTTPASTIPGPRGSGPELTFDPLPEP
AAPAGRSASRGHRKRRVLYPRVRQLPVEEPNPAKRLFLLLLTIVFCQILMAE
EGVPAPLPEDAPNAASLAPTPSPVLEPFLNLTSEPSYALDLSTFLQHPAAF"
172
conflict
/gene="dif-2"
/citation={3}
/replace="g"
328
/gene="dif-2"
/citation={3}
/replace="g"
483..>1230
/gene="dif-2"
673..675
/gene="dif-2"
/citation={3}
/replace="ag"
687..689
/gene="dif-2"
/citation={3}
/replace="tg"
783..785
/gene="dif-2"
/citation={3}
/replace="tg"
1020..1021
/gene="dif-2"
/citation={3}
/replace="att"
1120
/gene="dif-2"
/citation={3}
/replace="g"
1133
/gene="dif-2"
/citation={3}
/replace="t"
BASE COUNT 259 a 367 c 342 g 262 t
ORIGIN

Query Match 100.0%; Score 258; DB 53; Length 1230;
Best Local Similarity 100.0%; Pred. No. 6.8e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCCAGTCAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
DB 222 GTCCGGCGCCAGCTGCCAGTCAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 281

QY 61 CTCACCATCGTCTTCTGCGAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCCT 120
DB 282 CTCACCATCGTCTTCTGCGAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCCT 341

QY 121 CCAGAGGAGCCCTTAACGCCCATCCCTGGCGCCACCCCTGCTGCCCGCTCCGCGAG 180
DB 342 CCAGAGGAGCCCTTAACGCCCATCCCTGGCGCCACCCCTGCTGCCCGCTCCGCGAG 401

QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCTCCAG 240
DB 402 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCTCCAG 461
```

```
QY 241 CAACACCCGGCGCCCTTC 258
DB 462 CAACACCCGGCGCCCTTC 479

RESULT 3
AF039067 1309 bp mRNA PRI 21-AUG-1998
LOCUS Homo sapiens anti-death protein (IEX-1L) mRNA, complete cds.
DEFINITION AF039067
ACCESSION AF039067.1 GI:3449375
VERSION AF039067.1
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1309)
AUTHORS Wu,M.X., Ao,Z., Prasad,K.V.S., Wu,R. and Schlossman,S.F.
TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell
survival
JOURNAL Science 281 (5379), 998-1001 (1998)
MEDLINE 98369175
REFERENCE 2 (bases 1 to 1309)
AUTHORS Ao,Z. and Wu,M.X.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Tumor Immunology, Dana Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
Source
1..1309
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1309
/gene="IEX-1L"
1..582
/gene="IEX-1L"
/note="type II membrane protein"
/codon_start=1
/product="anti-death protein"
/protein_id="AAC32558.1"
/db_xref="GI:3449376"
/translation="MCHSRCHPTMTILOAPTTPASTIPGPRGSGPELTFDPLPEP
AAPAGRSASRGHRKRRVLYPRVRQLPVEEPNPAKRLFLLLTIVFCQILMAE
EGVPAPLPEDAPNAASLAPTPSPVLEPFLNLTSEPSYALDLSTFLQHPAAF"
BASE COUNT 262 a 393 c 369 g 285 t
ORIGIN

Query Match 100.0%; Score 258; DB 10; Length 1309;
Best Local Similarity 100.0%; Pred. No. 6.7e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCCAGTCAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
DB 322 GTCCGGCGCCAGCTGCCAGTCAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 381

QY 61 CTCACCATCGTCTTCTGCGAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCCT 120
DB 382 CTCACCATCGTCTTCTGCGAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCCT 441

QY 121 CCAGAGGAGCCCTTAACGCCCATCCCTGGCGCCACCCCTGCTGCCCGCTCCGCGAG 180
DB 442 CCAGAGGAGCCCTTAACGCCCATCCCTGGCGCCACCCCTGCTGCCCGCTCCGCGAG 501

QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCTCCAG 240
DB 502 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCTCCAG 561

QY 241 CAACACCCGGCGCCCTTC 258
DB 562 CAACACCCGGCGCCCTTC 579
```

```

c 22 39.4 15.3 188907 65 AC026986
c 23 39.2 15.2 150152 74 AL355502
c 24 39 15.1 144798 6 AC026758
c 25 38.4 14.9 10165 1 AE002089
c 26 38.2 14.8 4391 52 HSAJ4832
c 27 38.2 14.8 8391 29 MBU36763
c 28 38.2 14.8 38390 29 SC2H12
c 29 38.2 14.8 174044 69 AC069521
c 30 38 14.7 580 11 MMU80602
c 31 38 14.7 64990 51 AP000251
c 32 38 14.7 100000 34 AP000030
c 33 38 14.7 100000 34 AP000134
c 34 38 14.7 100000 51 AP000212
c 35 38 14.7 340000 51 AP001711
c 36 37.8 14.7 135119 56 AC011578
c 37 37.6 14.6 169292 70 AC074092
c 38 37.6 14.6 196978 58 AC015872
c 39 37.4 14.5 65671 71 AC083783
c 40 37.2 14.4 843 77 CNS01K6M
c 41 37.2 14.4 2043 11 AF102818
c 42 37.2 14.4 12664 1 AF005029
c 43 37.2 14.4 50502 52 HSA494016
c 44 37.2 14.4 69354 69 AC069342
c 45 37 14.3 576 77 CNS01J02

```

ALIGNMENTS

```

RESULT 1
AF083421 477 bp mRNA PRI 02-SEP-1998
LOCUS Homo sapiens radiation-inducible immediate early response gene IEX1
DEFINITION (IEX1) mRNA, complete cds.
ACCESSION AF083421
VERSION AF083421.1 GI:3511288
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS Kondratyev,A.D., Chung,K.N. and Jung,M.O.
TITLE Identification and characterization of a radiation-inducible
glycosylated human early-response gene
JOURNAL Cancer Res. 56 (7), 1498-1502 (1996)
MEDLINE 96181295
AUTHORS Kumar,R., Kobayashi,T., Warner,G.M., Wu,Y., Salisbury,J.L.,
Lingle,W. and Pittelkow,M.R.
TITLE A Novel Immediate Early Response Gene, IEX-1 Is Induced by
Ultraviolet Radiation in Human Keratinocytes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 477)
AUTHORS Kumar,R., Pittelkow,M.R. and Warner,G.M.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1998) Nephrology Research, Mayo Clinic, 200 1st
Street SW, Rochester, MN 55905, USA
FEATURES
source
1..477
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="keratinocyte; fetal osteoblast"
1..477
/gene="IEX1"
7..477
/gene="IEX1"
/feature="UV irradiation induced gene in keratinocytes"
/codon_start=1
/product="radiation-inducible immediate early response
gene IEX1"
/protein_id="AAC33793.1"
/db_xref="GI:3511288"

```

```

/translation="MCHSRSHPTMTLQAPHPAPSTIPGPRKSGSPELFTFDPLPEP
AAPAGRPASGRHRRKRVLYPRVRQLPVEEPNPAKRLFLILLTVFQILMAE
EGVPAPLPEDAPNAASLAPTVPSPVLEFNLTSSESDYALDLSLTLQOHPAAF"
167
/gene="IEX1"
/citation=[1]
/replace="g"
323
/gene="IEX1"
/citation=[1]
/replace="g"
BASE COUNT 72 a 205 c 117 g 83 t
ORIGIN
Query Match 100.0%; Score 258; DB 10; Length 477;
Best Local Similarity 100.0%; Pred. No. 8.3e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Caps 0;
Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
Db 217 GTCCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 276
Qy 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCCCTTCCCT 120
Db 277 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCCCTTCCCT 336
Qy 121 CCAGAGGAGCCCTTAACGGCGCATCCCTGGCGCCACCCCTGTGTCCTCCCGTCTCGAG 180
Db 337 CCAGAGGAGCCCTTAACGGCGCATCCCTGGCGCCACCCCTGTGTCCTCCCGTCTCGAG 396
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACATACCTCTGGACCTCAGACATTTCTCCAG 240
Db 397 CCCTTTAATCTGACTTCGGAGCCCTCGGACATACCTCTGGACCTCAGACATTTCTCCAG 456
Qy 241 CAACACCGCGCCGCTTC 258
Db 457 CAACACCGCGCCGCTTC 474
RESULT 2
HSDIF2 1230 bp mRNA PKI 12-AUG-1997
LOCUS Homo sapiens mRNA for DIF-2 protein.
DEFINITION HSDIF2
ACCESSION Y14551
VERSION Y14551.1 GI:2660550
KEYWORDS dif-2 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1230)
AUTHORS Pietzsch,A., Buchler,C., Aslanidis,C. and Schmitz,G.
TITLE Identification and characterization of a novel monocyte/macrophage
differentiation-dependent gene that is responsive to
lipopolysaccharide, ceramide, and lysophosphatidylcholine
JOURNAL Biochem. Biophys. Res. Commun. 235 (1), 4-9 (1997)
MEDLINE 97339426
REFERENCE 2 (bases 1 to 1230)
AUTHORS Pietzsch,A.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1997) A. Pietzsch, Institute for Clinical
Chemistry and Laboratory Medicine, University of Regensburg,
Franz-Josef-Strauss-Allee 11, 93053 Regensburg, FRG
REFERENCE 3 (bases 1 to 1230)
AUTHORS Kondratyev,A.D., Chung,K.N. and Jung,M.O.
TITLE Identification and characterization of a radiation-inducible
glycosylated human early-response gene
JOURNAL Cancer Res. 56 (7), 1498-1502 (1996)
MEDLINE 96181295
FEATURES
Location/Qualifiers
1..1230
/organism="Homo sapiens"

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 30, 2001, 19:23:39 ; Search time 1900.13 Seconds
(without alignments)
694.889 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468
Perfect score: 258
Sequence: 1 CTCGGCGCCGCTGCCACT.....AGCAACACCGCGCGCGCTTC 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_ph.*
- 6: gb_pl1.*
- 7: gb_pl2.*
- 8: gb_pri.*
- 9: gb_pr2.*
- 10: gb_pr3.*
- 11: gb_ro.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: em_fun.*
- 15: em_hum1.*
- 16: em_hum2.*
- 17: em_in.*
- 18: em_om.*
- 19: em_or.*
- 20: em_ov.*
- 21: em_pat.*
- 22: em_ph.*
- 23: em_pl.*
- 24: em_ro.*
- 25: em_sts.*
- 26: em_sy.*
- 27: em_un.*
- 28: em_v1.*
- 29: gb_ba3.*
- 30: gb_in1.*
- 31: gb_in2.*
- 32: gb_in3.*
- 33: gb_pl3.*
- 34: gb_pr4.*
- 35: em_bal.*
- 36: em_ba2.*
- 37: em_htg1.*
- 38: em_htg2.*
- 39: em_htg3.*
- 40: em_htg4.*
- 41: em_htg5.*
- 42: em_htg6.*
- 43: em_htg7.*

- 44: em_htg8.*
- 45: em_htg9.*
- 46: em_htg10.*
- 47: em_hum3.*
- 48: em_hum4.*
- 49: em_hum5.*
- 50: em_hum6.*
- 51: gb_pr5.*
- 52: gb_pr6.*
- 53: gb_pr7.*
- 54: gb_hg1.*
- 55: gb_hg2.*
- 56: gb_hg3.*
- 57: gb_hg4.*
- 58: gb_hg5.*
- 59: gb_hg6.*
- 60: gb_hg7.*
- 61: gb_hg8.*
- 62: gb_hg9.*
- 63: gb_hg10.*
- 64: gb_hg11.*
- 65: gb_hg12.*
- 66: gb_hg13.*
- 67: gb_hg14.*
- 68: gb_hg15.*
- 69: gb_hg16.*
- 70: gb_hg17.*
- 71: gb_hg18.*
- 72: gb_hg19.*
- 73: gb_hg20.*
- 74: gb_hg21.*
- 75: gb_hg22.*
- 76: gb_hg23.*
- 77: gb_sts1.*
- 78: gb_sts2.*
- 79: gb_vil.*
- 80: gb_v12.*
- 81: gb_pat1.*
- 82: gb_pat2.*
- 83: em_hg0.*
- 84: gb_htg24.*
- 85: gb_pr8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	258	100.0	477	10	AF083421	AF083421 Homo sapi
2	258	100.0	1230	53	HSD1F2	Y14551 Homo sapien
3	258	100.0	1309	10	AF039067	AF039067 Homo sapi
4	258	100.0	1693	10	AF071596	AF071596 Homo sapi
5	258	100.0	1864	53	HSPRG1	X96438 H.sapiens p
6	258	100.0	44118	9	AC006165	AC006165 Homo sapi
7	258	100.0	90244	8	AB023051	AB023051 Homo sapi
8	258	100.0	200000	51	AP000512	AP000512 Homo sapi
9	256.4	99.4	1223	85	S81914	S81914 TEX-1-radia
10	154.4	59.8	1758	11	RNPRG1	X96437 R.norvegicu
11	151.2	58.6	203476	70	AC074150	AC074150 Mus muscu
12	149.6	58.0	1938	11	MMGLY96	X67644 M.musculus
13	139.2	54.0	187868	62	AC022301	AC022301 Mus muscu
14	44	17.1	297	81	A75448	A75448 Sequence 11
15	44	17.1	297	81	A78427	A78427 Sequence 11
16	43.2	16.7	115468	57	AC011820	AC011820 Homo sapi
17	43	16.7	7218	81	I66494	I66494 Sequence 14
18	40.2	15.6	34725	29	SC1B2	AL356812 Streptomy
c 19	39.8	15.4	3897	2	MXU40656	U40656 Myxococcus
c 20	39.6	15.3	33818	2	MTCY159	Z83863 Mycobacteri
21	39.4	15.3	2003	85	HUMDAFC1	M64356 Human decay

```

BASE COUNT      66 a      117 c      61 g      52 t      1 others
ORIGIN

Query Match      12.3%; Score 151; DB 81; Length 297;
Best Local Similarity 76.0%; Pred No. 1.6e-24;
Matches 212; Conservative 0; Mismatches 65; Indels 2; Gaps 2;

Qy 1 ATGTGTCACCTCGCAGCTGCCACCGACCATGACCATCTGTCAGGCCCCGACCCGGCC 60
   ||| ||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||
Db 16 ATGGAACACTCTCGCGGTGCCACTCCCCCATGACCACTTACAGGCCCTAACCCAGCC 75
   ||| ||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||
Qy 61 CCTCCACCATCCCGGACCCCGGGGGCTCCGGTCTGAGATCTTCACCTTCGACCT 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 CCTCAGCTTATCGTCCAGCCCAATA-GCTGCCGCCCTGAGATCTCCACCTCAGACCT 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 121 CTCCGGAGCCCGAGGGCCCTGCCGGGGCCCGCCAGCCCTCTCGGGGCACCGAAAG 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 CTCCGGAGCCCGAGGGCCCTGCCGGGGCCCGCCAGCTTC-CTCGAGGCCACCAAG 193
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 CGCAGCCCGAGGGTTCTTACCTCGAGTGTCCGGGCCAGTGCAGTCGAGGAACCG 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 194 CGCAGTTCCAGGAATCTCTACCTAGAGTATCCGGGCCAGTTGCCAGTTAAGGACCA 253
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 AACCCAGCCAAAGGCTTCTCTTCTTCTGCTGCTCACCAATC 279
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 254 AACCTGCCAGAGGCTTCTTTTCGGGTTGATCGCCATC 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: January 30, 2001, 19:23:39
Job time: 25905 sec


```

Oy 312 TGTGCGCGGCCCTGCTCCAGAGACGCCCTTAAGCGCGCATCC-----T 359
Db 1132 TGTGTCGACGCCCTGCTCCGAGGATGTACACGCCGCTGACACCTGAGCCCATTC 1191
Oy 360 GGGCGCCACCCCTGTGTGCCCGCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGA 419
Db 1192 TGGCCCATTAAGTGTGCGGCCCGCTCGAGCCCTTTGAACCTGACCTCGGAGCTCTCGGA 1251
Oy 420 CTAGGCTCTGGACTGACACTTCTCTCCAGCAACACCCCGCGCTTCTAACTGTGACT 479
Db 1252 CTAGGCTCTGGACTTAAAGCTTTCTCTAGCAACATCCGCGCGCTTCTAAAGCGATG 1311
Oy 480 CCGCGCACTCCCAAAAGAAATCGGAAAACCAACAAACACCGGCTGACTGTGTGC 539
Db 1312 GCTC-----ACAATCGGAAGAAACAAAGCCACCATGGATGGTACTGTGTGC 1358
Oy 540 GCGAGACGATATCCCAACTGGGACTTCCGAGGCAACTTGAATCTGAAACACTACAGCG 599
Db 1359 GAGAGACGATATCCCAACTGGGATTTCTTAAGCAAGCTTAATCTGAAACACTACGCCA 1418
Oy 600 AGACGCCACCCGGTGTG 618
Db 1419 AGACACACCGCGGTCTG 1437

RESULT 14
AC022301 LOCUS
DEFINITION Mus musculus clone RP23-329K24, *** SEQUENCING IN PROGRESS ***, 59
unordered pieces.
AC022301
VERSION AC022301.9 GI:8493517
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 187868)
Metzker,M.L., Lewis,S., Okwuonu,G., Carlock,C., Garner,T.,
Dederich,D., Thomas,S., Okwuonu,G., Bonnini,D., Brooks,A., Brown,J.,
Addison,S., Pace,A., Williams,G., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Buhay,C., Bunac,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Cox,C., Davis,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Fernandez,C., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
Gorell,J.H., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Hosak,H., Hou,X., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 187868)
Worley,K.C.
Direct Submission
Submitted (29-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 13, 2000 this sequence version replaced gi:8248590.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAAH
Center clone name: RP23-329K24
----- Summary Statistics
Sequencing vector: M13; L08821

```

```

Chemistry: Dye-primer Bodipy: 82% of reads
Chemistry: Dye-terminator Big Dye: 18% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 13377 bases at least Q40
Consensus quality: 159061 bases at least Q30
Consensus quality: 169688 bases at least Q20
Estimated insert size: 166938; sum-of-contigs estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7154: contig of 7154 bp in length
* 7155 7254: gap of unknown length
* 7255 14340: contig of 7086 bp in length
* 14341 14440: gap of unknown length
* 14441 22916: contig of 8476 bp in length
* 22917 23016: gap of unknown length
* 23017 30834: contig of 7818 bp in length
* 30835 30934: gap of unknown length
* 30935 37933: contig of 7059 bp in length
* 37934 38093: gap of unknown length
* 38094 45511: contig of 7418 bp in length
* 45512 51842: contig of 6231 bp in length
* 51843 51942: gap of unknown length
* 51943 58572: contig of 6630 bp in length
* 58573 58672: gap of unknown length
* 58673 62719: contig of 4046 bp in length
* 62719 62819: gap of unknown length
* 62819 67974: contig of 5156 bp in length
* 67975 68075: gap of unknown length
* 68075 72883: contig of 4809 bp in length
* 72884 72984: contig of 3743 bp in length
* 72984 76727: gap of unknown length
* 76727 79631: contig of 2805 bp in length
* 79632 79731: gap of unknown length
* 79732 82858: contig of 3127 bp in length
* 82859 82958: gap of unknown length
* 82959 86054: contig of 3096 bp in length
* 86055 86154: gap of unknown length
* 86155 90283: contig of 4129 bp in length
* 90284 90383: gap of unknown length
* 90384 94598: contig of 4215 bp in length
* 94599 94698: gap of unknown length
* 94699 99014: contig of 4316 bp in length
* 99015 103296: contig of 4181 bp in length
* 103296 103396: gap of unknown length
* 103396 106699: contig of 3303 bp in length
* 106699 106799: gap of unknown length
* 106799 109658: contig of 2859 bp in length
* 109658 109757: gap of unknown length
* 109758 113392: contig of 3635 bp in length
* 113393 113492: gap of unknown length
* 113493 117011: contig of 3519 bp in length
* 117012 120036: contig of 2925 bp in length
* 120037 120136: gap of unknown length
* 120137 122449: contig of 2313 bp in length
* 122450 122549: gap of unknown length
* 122550 124276: contig of 1726 bp in length
* 124276 124376: gap of unknown length
* 124376 126297: contig of 1921 bp in length
* 126297 126397: gap of unknown length
* 126397 129180: contig of 2784 bp in length
* 129181 129280: gap of unknown length

```



```
Db 67 CCTCCACCATCCGGGAGCCCGGGGGCTCCGGTCCCTGAGATCTTCACCTTCGACCCCT 126
Qy 121 CTCCTGGAGCCGCGAGCGCCCTCCCGGGCGCCCGAGGCTCTCGGGGCGACCGAAG 180
Db 127 CTCCTGGAGCCGCGAGCGCCCTCCCGGGCGCCCGAGGCTCTCGGGGCGACCGAAG 186
Qy 181 CCGAGCGGAGGCTTCTCTACCTCGAGTGGTCCGGCGCCAGCTGCCAGTCGAGAACCG 240
Db 187 CCGAGCGGAGGCTTCTCTACCTCGAGTGGTCCGGCGCCAGCTGCCAGTCGAGAACCG 246
Qy 241 AACCCAGCAAAAGGCTTCTCTTGTGTGTGTCACCATGCTTCTGCGCAGATCTCTGATG 300
Db 247 AACCCAGCAAAAGGCTTCTCTTGTGTGTGTCACCATGCTTCTGCGCAGATCTCTGATG 306
Qy 301 GCTGAGAGGGTGTCCGGCGCCCTGCTCCAGAGGAGCCCTTAACGCCGATCCCTG 360
Db 307 GCTGAGAGGGTGTCCGGCGCCCTGCTCCAGAGGAGCCCTTAACGCCGATCCCTG 366
Qy 361 GCGCCACCCCTGTGTCCCGCTGCTCGAGCCCTTAACTGACATTCGGAGCCCTCGGAC 420
Db 367 GCGCCACCCCTGTGTCCCGCTGCTCGAGCCCTTAACTGACATTCGGAGCCCTCGGAC 426
Qy 421 TAGCCTCTGAGACTAGCACTTCTCTCCAGCAACCCCGCGCTTCTAA 471
Db 427 TAGCCTCTGAGACTAGCACTTCTCTCCAGCAACCCCGCGCTTCTAA 477

RESULT 10
HSA227914
LOCUS Homo sapiens partial mRNA; ID YG40-2. PRI 17-DEC-1998
DEFINITION AJ227914
ACCESSION AJ227914.1 GI:3183967
VERSION AJ227914.1
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS Dominguez,O.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1998) Dominguez O., Immunologia, Hospital Germans
Trias i Pujol, 08916 Badalona, and Research Center, Almirall
Prodesfarma, 08024 Barcelona, Spain
REFERENCE 2 (bases 1 to 343)
AUTHORS Dominguez,O., Ashhab,Y., Sabater,L., Belloso,E., Caro,P. and
Pujol-Borrell,R.
TITLE Cloning of ARE-containing genes by AU-motif-directed display
JOURNAL Genomics 54 (2), 278-286 (1998)
MEDLINE 99047534
COMMENT Overlapping sequences: Y14551.
FEATURES
source
1..343
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="adult"
/cell_type="T lymphocyte"
/cell_line="T12.12"
/clone="PH19-20, PH19-21"
<1..>343
/gene="DIF-2"
1..343
/gene="DIF-2"

mRNA
72 a 67 c 114 g 85 t 5 others
gene
BASE COUNT 72 a 67 c 114 g 85 t 5 others
ORIGIN
Query Match 24.3%; Score 298.8; DB 52; Length 343;
Best Local Similarity 97.1%; Pred. No. 1.4e-58;
Matches 333; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Qy 862 GCTTTGCGGACTGGTGAGAGCAGGACCTCGAACTCGGCGCAAGTAGGAGAGAAATG 921
|||||
```

```
Db 1 GCTTTGCGGACTGGTGAGAGCAGGAC-CCITGAACCTCCGCAAGTAGGAGAAG-AAATG 58
Qy 922 GGGAGGACTCGGCTGGGGAGGAGAGTCCCGGCTGGGATGAAGTCTGTGGTGGGTCTGTA 981
Db 59 GGGAGGACTCGGCTGGGGAGGAGAGTCCCGGCTGGGATGAAGTCTGTGGTGGGTCTGTA 118
Qy 982 GTTTAGGAGGTGACTTGCATCTCCAGCATCTCAACTCCGCTGTCTACTGTCTGAGACTT 1041
Db 1119 G-TTAGGAGGTGACTTGCATCTCCAGCATCTCAACTCCGCTGTCTACTGTCTGAGACTT 177
Qy 1042 CGGGGAGCCATTAGGATGAGATCCGTGAGATCTTCCATCTTCTTGAAGTCGGCTTTAG 1101
Db 178 CGGGGAGCCATTAGGATGAGATCCGTGAGATCTTCCATCTTCTTGAAGTCGGCTTTAN 237
Qy 1102 GGTGCTCGGAGGTGAGGGTTGGGGTTGGTGGCTCTCACGGAGCGACTGTCGAGATC 1161
Db 238 GGTGCTCGGAGGTGAGGGTTGGGGTTGGTGGCTCTCACGGAGCGACTGTCGAGATC 297
Qy 1162 GCCTAGTATGTTCTGTGAACACAAATAAAATTGATTTACTGTC 1204
Db 298 GCCTANTATGTTCTGTGAACACAAATAAAATTGATTTACTGTC 340

RESULT 11
RNPRG1
LOCUS R.norvegicus PRG1 gene. ROD 03-SEP-1996
DEFINITION X96437
ACCESSION X96437.1 GI:1515318
VERSION X96437.1
KEYWORDS PRG1 gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 1758)
AUTHORS Schaffer,H., Trauzold,A., Siegel,E.G., Polsch,U.R. and Schmidt,W.E.
TITLE PRG1: a novel early-response gene transcriptionally induced by
pituitary adenylate cyclase activating polypeptide in a pancreatic
carcinoma cell line
JOURNAL Cancer Res. 56 (11), 2641-2648 (1996)
MEDLINE 96221139
REFERENCE 2 (bases 1 to 1758)
AUTHORS Trauzold,A.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1996) A. Trauzold, Laboratory of Molecular
Gastroenterology, 1st Dept. of Medicine, University of Kiel,
Schittenhelmstrasse 12, Kiel, D-24105, FRG
FEATURES
source
1..1758
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/cell_line="AR4-2J"
590..1279
/gene="PRG1"
590..1279
/gene="PRG1"

BASE COUNT 358 a 475 c 470 g 455 t
ORIGIN
Query Match 19.7%; Score 242.4; DB 11; Length 1758;
Best Local Similarity 62.1%; Pred. No. 1.3e-45;
Matches 589; Conservative 0; Mismatches 181; Indels 178; Gaps 6;

Qy 1 ATGTGTCACTCTCGAGCTGCCACCGACCATGACCATCTCGAGCCCGCCCGCCCGCC 60
|||||
Db 590 ATGTGCGATTCCGCTGAACACCTCCACACCATGACTGGCCGAGGGCTCCGGCCAGTT 649
Qy 61 CCTCCACCATCCCGGAGCCCGGGGCTCCGGTCTCTGAGATCTTCACTTCACTTCACT 120
|||||
Db 650 CCTCCACCGCCCGGAGCACTCCGGTCCCGGGCTCCGGTCCCGGAAATTTTCACTTCACT 709
|||||
Qy 121 CTCCCGGAGCCCGCAGCGCCCTCTCGGGGGCCCGCCCTCTCGGGGGCCCGCCAGAAC 180
|||||
```


Qy	250	AAAAGCGCTCTCTTTCTGCTGCTCACCATCCTCTTCTCCAGATCCTGATGGCTGAAGAG	309
Db	361	AAAAGGCTTCTCTTTCTGCTGCTCACCATCCTCTTCTGCCAGATCCTGATGGCTGAAGAG	420
Qy	310	GGTGTGGCGGCGCCCTGCTCCAGAGAGCGCCCTAACCGCGGATCCCTGGGCGCCACC	369
Db	421	GGTGTGGCGGCGCCCTGCTCCAGAGAGCGCCCTAACCGCGGATCCCTGGGCGCCACC	480
Qy	370	CCTGTGTCGCCCGCTCGAGCGCTTAAATCTGACTTCGGAGCCCTCGGACTAGCTCTG	429
Db	481	CCTGTGTCGCCCGCTCGAGCGCTTAAATCTGACTTCGGAGCCCTCGGACTAGCTCTG	540
Qy	430	GACCTCAGCACTTTCTCCAGCAACACCCCGCGCCCTTCTAACTGTACATCCCGGCACTC	489
Db	541	GACCTCAGCACTTTCTCCAGCAACACCCCGCGCCCTTCTAACTGTACATCCCGGCACTC	600
Qy	490	CCAAAAGAATCGGAAACACCAAGAAACACACCGGCGCATAGAGACCGGAGCAGCT	549
Db	601	CCAAAAGAATCGGAAACACCAAGAAACACACCGGCGCATAGAGACCGGAGCAGCT	660
Qy	550	ATCCCCAATGGGACTTCCGAGGCACTTGAACCTCAGAACACTACACGCGGAGCGCCACC	609
Db	661	ATCCCCAATGGGACTTCCGAGGCACTTGAACCTCAGAACACTACACGCGGAGCGCCACC	720
Qy	610	CGGTGCTTGAGCGCGGACCGAGCGCACAGAGACCGGCGCATAGAGACCGGAGCAGAC	669
Db	721	CGGTGCTTGAGCGCGGACCGAGCGCACAGAGACCGGCGCATAGAGACCGGAGCAGAC	780
Qy	670	CCCAGCTGGGGCTAGGCCCGGTTGGGAAGAGAGCGTCGTTAAATTTATTTCTTATGCTCC	729
Db	781	CCCAGCTGGG--CTAGGCCGCTGGGAAGAGAGCGTCCTTAAATTTATTTCTTATGCTCC	838
Qy	730	TAATTAATATTTATATCTATTTATGTACGTCCTCTAGTGTATGGAGATGTGTACGTAAT	789
Db	839	TAATTAATATTTATATCTATTTATGTACGTCCTCTAGTGTATGGAGATGTGTACGTAAT	898
Qy	790	ATTTATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTGTAATTCAGGCTCTCTT	849
Db	899	ATTTATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTGTAATTCAGGCTCTCTT	958
Qy	850	GGTATTTATTTGAGCTTTGTGGGACTGTGTGGAAGCAGACACCTCGGACTGGGATGAAGTCTG	909
Db	959	CGTATTTATTTGAGCTTTGTGGGACTGTGTGGAAGCAGACACCTCGGACTGGGATGAAGTCTG	1018
Qy	910	GGAGAGAAATGGGAGGACTCGGGTGGGAGGAGCGTCCCGGCTGGGATGAAGTCTG	969
Db	1019	GGAGAGAAATGGGAGGACTCGGGTGGGAGGAGCGTCCCGGCTGGGATGAAGTCTG	1078
Qy	970	GGTGGGTCGTAACTTTAGGAGGTGACTTGCATCCTCCAGCATCTCAACTCCGCTGTGCTAC	1029
Db	1079	GGTGGGTCGTAACTTTAGGAGGTGACTTGCATCCTCCAGCATCTCAACTCCGCTGTGCTAC	1138
Qy	1030	TGTGTGAGACTTCGGGGAGCATTAGGAATGAGATCCGTGAGATCTTCCATCTTCTTGA	1089
Db	1139	TGTGTGAGACTTCGGGGAGCATTAGGAATGAGATCCGTGAGATCTTCCATCTTCTTGA	1198
Qy	1090	AGTCGGCTTTAGGCTGGCTCGAGGTAGAGGGTTGGGGTTGTGGGCTGTACGAGAGCG	1149
Db	1199	AGTCGGCTTTAGGCTGGCTCGAGGTAGAGGGTTGGGGTTGTGGGCTGTACGAGAGCG	1258
Qy	1150	ACTGTCGAGATCCCTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC	1200
Db	1259	ACTGTCGAGATCCCTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC	1309
RESULT	7		
AF071596			
LOCUS	1693 bp	DNA	10-NOV-1998
DEFINITION	Homo sapiens apoptosis inhibitor (IEX-1L) gene, complete cds.		
ACCESSION	AF071596		
VERSION	AF071596.1 GI:3851531		
KEYWORDS	human.		
SOURCE			

ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Wu, M.X., Ao, Z., Prasad, K.V., Wu, R. and Schlossman, S.F.		
TITLE	IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival		
JOURNAL	Science 281 (5379), 998-1001 (1998)		
MEDLINE	98369175		
REFERENCE	2 (bases 1 to 1693)		
AUTHORS	Wu, M.X. and Ao, Z.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-JUN-1998) Tumor Immunology, Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA		
FEATURES	Location/Qualifiers		
source	1..1693		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
mRNA	359..1693		
	/gene="IEX-1L"		
gene	/product="apoptosis inhibitor"		
	359..1693		
	/gene="IEX-1L"		
CDS	385..966		
	/gene="IEX-1L"		
	/note="immediate early response protein"		
	/codon_start=1		
	/product="apoptosis inhibitor"		
	/protein_id="AAC72344.1"		
	/db_xref="GI:3851532"		
	/translation="MCHSRSCYPTMTLQAPTPAPSTIPGRKSGPSELTFTFPIPLPAP		
	AAAPGRPSASGRHKRRSRVSLPSSGEYRRSGHSCALPWSHWGTIRLOSLLDITCII.		
	FCPLVRLPVEEPNPARKLLFLITVFCQLMAEGVPAPLPPEPADAPNAASLAPT		
	PVSPVLEPENLTSEPSYDALDLSLFLQHPAAF"		
BASE COUNT	348 a	529 c	456 g
ORIGIN	360 t		
Query Match	85.4%	Score 1048.6;	DB 10; Length 1693;
Best Local Similarity	91.1%;	Pred. No. 2.6e-231;	
Matches 1194;	Conservative	0; Mismatches	4; Indels 113; Gaps 3;
Qy	1	ATGTGTACATCTCGAGCTGCCACCGGACCATCACCATCTGCGAGGCCCGCCGCCCC	60
Db	385	ATGTGTACATCTCGAGCTGCCACCGGACCATCACCATCTGCGAGGCCCGCCGCCCC	444
Qy	61	CCCTCCACCATCCCGGGACCCCGGGGCTCCGCTCGATCTTACCTTCGACCTT	120
Db	445	CCCTCCACCATCCCGGGACCCCGGGGCTCCGCTCGATCTTACCTTCGACCTT	504
Qy	121	CTCCGAGAGCCGACGGGCCCTGCGGGGCGCCCGACGCGCTCTCGGGGCGACCGAAG	180
Db	505	CTCCGAGAGCCGACGGGCCCTGCGGGGCGCCCGACGCGCTCTCGGGGCGACCGAAG	564
Qy	181	CGAGCGGAGGGT-TCTCTACCTCGAGTG	210
Db	565	CGAGCGGAGGGTCTCTTACCTCGAGTGAGTATGCGGAAAGTGGGATTCGCGG	624
Qy	211	-----	210
Db	625	TGTGCGCTGCCCTGGAGTCACATGGGGAACGACCCGACTCCAGAGCCCTCGACCTGACCTGT	684
Qy	211	-----GTCCGGGCGCCAGCTGCCAGTCAGTCGAGGAACCGAACCAGCC	249
Db	685	CTCCTGTTTGTCTCCCTTAGTCCGCGCCAGCTGCCAGTCGAGGAACCGAACCAGCC	744
Qy	250	AAAAGCTTCTCTTTCTGCTCACCATCGTCTTCTGCCAGATCTCTGATGGCTGAAGAG	309
Db	745	AAAAGCTTCTCTTTCTGCTCACCATCGTCTTCTGCCAGATCTCTGATGGCTGAAGAG	804
Qy	310	GGTGTGCGGCGCCCTGCTCCAGAGAGCGCCCTTAACGCCGCTCTCTGCGGCCACC	369
Db	805	GGTGTGCGGCGCCCTGCTCCAGAGAGCGCCCTTAACGCCGCTCTCTGCGGCCACC	864

Db 97843 CGCAGCCGACGGTTCTCTACCTCTGAGTGGTGGTATGAGTATGCCGGAAGTGGCATTCGCGGG 97902
Qy 210 ----- 209
Db 97903 CTCGCTGCCCTGGAGTCACTCGGGGAACGACCCGACTCCAGAGCCCTCGACCTGACCTGT 97962
Qy 210 -----GCTCGCGGCCAGCTGCCAGTGGAGAACCGAACCCAGC 248
Db 97963 CTCCTGTTTGTCTCCCTTAGGTCCGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGC 98022
Qy 249 CAAGAAGCTTCTCTTCTGCTGCTCACCATCTCTTCTGCCAGATCTGTATGCTGCTGAAGA 308
Db 98023 CAAGAAGCTTCTCTTCTGCTGCTCACCATCTCTTCTGCCAGATCTGTATGCTGCTGAAGA 98082
Qy 309 GGTGTGTCGCGCGCCCTGCTCCAGAGGACGCCCTTAAGCCGCGCATCCCTGCGCGCCAC 368
Db 98083 GGTGTGTCGCGCGCCCTGCTCCAGAGGACGCCCTTAAGCCGCGCATCCCTGCGCGCCAC 98142
Qy 369 CCCTGTGTCGCGCGCTCTCGAGGCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCT 428
Db 98143 CCCTGTGTCGCGCGCTCTCGAGGCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCT 98202
Qy 429 GGACTTCAGACATTTCTCTCCAGAACACCCGCGGCTTCTAATCTGACTTCGCTCCCGACT 488
Db 98203 GGACTTCAGACATTTCTCTCCAGAACACCCGCGGCTTCTAATCTGACTTCGCTCCCGACT 98262
Qy 489 CCCCAAAAGAAATCCGAAACCAACCAAGAAACACAGGCTTACCTGGTCCGCGAGCG 548
Db 98263 CCCCAAAAGAAATCCGAAACCAACCAAGAAACACAGGCTTACCTGGTCCGCGAGCG 98322
Qy 549 TATCCCAACTGGGACTTTCGAGGCACTTGAATCTGAACTCAGAACACTACAGCGGAGACGCCAC 608
Db 98323 TATCCCAACTGGGACTTTCGAGGCACTTGAATCTGAACTCAGAACACTACAGCGGAGACGCCAC 98382
Qy 609 CCGTGTCTCAGCGGGAGCGGACGACAGACACCGAGCGGATGAGGAGGAGACGCCAC 668
Db 98383 CCGTGTCTCAGCGGGAGCGGACGACAGACACCGAGCGGATGAGGAGGAGACGCCAC 98442
Qy 669 GCCAGCTGGGGCTAGCCCGGCTGGGAAGGAGAGCGCTGTTAATTTATTCTTATTGCTC 728
Db 98443 GCCAGCTGGGGCTAGCCCGGCTGGGAAGGAGAGCGCTGTTAATTTATTCTTATTGCTC 98502
Qy 729 CTAAATAATTTATATGATTTATGTAGCTCTCTAGTATGAGATGATGATGATGATGATGATGAT 788
Db 98503 CTAAATAATTTATATGATTTATGTAGCTCTCTAGTATGAGATGATGATGATGATGATGATGAT 98562
Qy 789 TATTTATTTTAACTTATGAAGGCTGTGAGATGTTCCCTCTGCTGTAATTCGAGGCTCT 848
Db 98563 TATTTATTTTAACTTATGAAGGCTGTGAGATGTTCCCTCTGCTGTAATTCGAGGCTCT 98622
Qy 849 TGGTATTTATGAGCTTTGTGGACTTGTGGAGAGGACACTTGGAACTTGGCGGCAAGT 908
Db 98623 TGGTATTTATGAGCTTTGTGGACTTGTGGAGAGGACACTTGGAACTTGGCGGCAAGT 98682
Qy 909 AGGAGAAGAAATGGGAGGACTCGGGTGGGGAGGAGCTCCCGGCTGGGATGAAGTCTGG 968
Db 98683 AGGAGAAGAAATGGGAGGACTCGGGTGGGGAGGAGCTCCCGGCTGGGATGAAGTCTGG 98742
Qy 969 TGGTGGGTCTGAATTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCGCTGCTGCTA 1028
Db 98743 TGGTGGGTCTGAATTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCGCTGCTGCTA 98802
Qy 1029 CTGTGTCAGACTTTCGCGGACCAATTAGGAATGAGATCCCTGAGATCTCTTCCATCTTCTTG 1088
Db 98803 CTGTGTCAGACTTTCGCGGACCAATTAGGAATGAGATCCCTGAGATCTCTTCCATCTTCTTG 98862
Qy 1089 AAGTCGCTTTAGGGTGGCTGCGAGGTAGAGGTTGGGGTGGTGGGCTGTACGGAGC 1148
Db 98863 AAGTCGCTTTAGGGTGGCTGCGAGGTAGAGGTTGGGGTGGTGGGCTGTACGGAGC 98922
Qy 1149 GACTGTGAGATCCGCTAGTATGTTCTGTAACACAAATTAATAATGATTACTGTC 1204
Db 98923 GACTGTGAGATCCGCTAGTATGTTCTGTAACACAAATTAATAATGATTACTGTC 98978

RESULT 6
AF039067 1309 bp mRNA PRI 21-AUG-1998
LOCUS Homo sapiens anti-death protein (IEX-1L) mRNA, complete cds.
DEFINITION AF039067
ACCESSION AF039067
VERSION AF039067.1 GI:3449375
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1309)
AUTHORS Wu.M.X., Ao.Z., Prasad.K.V.S., Wu.R. and Schlossman,S.F.
TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival
JOURNAL Science 281 (5379), 998-1001 (1998)
MEDLINE 98369175
REFERENCE 2 (bases 1 to 1309)
AUTHORS Ao.Z. and Wu.M.X.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Tumor Immunology, Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
Location/Qualifiers
1..1309
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1309
/gene="IEX-1L"
1..582
/gene="IEX-1L"
/note="type II membrane protein"
/codon_start=1
/product="anti-death protein"
/protein_id="AAC32558.1"
/db_xref="GI:3449376"
/translation="MCHSRSCHTPTMTILOAPTAPSTIPGRGSGPEIFTFDPLPEP
AAAPGSPASRGRKSRVSLPSSEYRSHSKALPWSHGTPRIQSLDTCLL
FCLPLVRQLPVEPNPAKRLILFVLCIILMAFEGVPAPLPPEDAPNAASLAFT
PVSPLRPFNLTSPSPDYALDSTFTQOHFAAF"
BASE COUNT 262 a 393 c 369 g 285 t
ORIGIN
Query Match 85.4%; Score 1048.6; DB 10; Length 1309;
Best Local Similarity 91.1%; Pred. No. 2.7e-231;
Matches 1194; Conservative 0; Mismatches 4; Indels 113; Gaps 3;
Qy 1 ATGTGTCACTCTCGAGCTGCCACCCGACCATGACCATCTCGAGCGCCCGACCCCGGCC 60
Db 1 ATGTGTCACTCTCGAGCTGCCACCCGACCATGACCATCTCGAGCGCCCGACCCCGGCC 60
Qy 61 CCCTCCACATCCCGGGACCCCGGGGCTCCGGTCTCTGAGATCTTCACCTTCGACCT 120
Db 61 CCCTCCACATCCCGGGACCCCGGGGCTCCGGTCTCTGAGATCTTCACCTTCGACCT 120
Qy 121 CTCGGGAGCCGCGAGCGCCCTCCCGGGCGCCCGCCAGCGCTCTCGCGGGACCGGAAAG 180
Db 121 CTCGGGAGCCGCGAGCGCCCTCCCGGGCGCCCGCCAGCGCTCTCGCGGGACCGGAAAG 180
Qy 181 CGAGCCGACGGGT-TCTCTACCTCGAGTG----- 210
Db 181 CGAGCCGACGGGTCTCTCTACCTTCGAGTGGTGGTATTCGCCCAAGTGGGCAATTTCGGG 240
Qy 211 ----- 210
Db 241 TGTGCTGCCCTGGAGTCACTGGGGAACGACCCGACTCCAGAGCTCGACCTGACCTGT 300
Qy 211 -----GTCCGGCCGCGCTGCCAGTCCAGTCCGAGGACCGAACCCAGCC 249
Db 301 CTCCTGTTTGTCTCCCTTTAGTCCGGCGCAGCTGCCAGTCCGAGGAACCGAACCCAGCC 360

```
Db 61518 CTAATTAATATTATATGTAATTTATGTAAGTCTCTAGGTGATGGAGATGTGTACGTAA 61577
Qy 789 TATTTTATTTTAACTTATCCAAAGGTGTCAGATGTTCCCTCTGCTGCTAAATCAGGTCCTCT 848
Db 61578 TATTTTATTTTAACTTATCCAAAGGTGTCAGATGTTCCCTCTGCTGCTAAATCAGGTCCTCT 61637
Qy 849 TGGTATTTATTTAGCTTTGTGGGACTGCTGGAAGCAGGACACCTGGAACTCGCGCAAGT 908
Db 61638 TCGTATTTTATTTAGCTTTGTGGGACTGCTGGAAGCAGGACACCTGGAACTCGCGCAAGT 61697
Qy 909 AGGAGAGAAATGGGGAGACTTCGGGTGGGGAGAGACCTCCAGCATCTCAACTCGCTGTGCTA 1028
Db 61698 AGGAGAGAAATGGGGAGACTTCGGGTGGGGAGAGACCTCCAGCATCTCAACTCGCTGTGCTA 61757
Qy 969 TGGTGGCTCTTAAGTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCGCTGTGCTA 1028
Db 61758 TGGTGGGTGGTAACTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCGCTGTGCTA 61817
Qy 1029 CTGTGTAGACTTCGGCGGACCATTAGGAATGAGATCGTGAGATCGTTCATCTCTTCTG 1088
Db 61818 CTGTGTAGACTTCGGCGGACCATTAGGAATGAGATCGTGAGATCGTTCATCTCTTCTG 61877
Qy 1089 AAGTCGCCCTTTAGGTTGCTCGGAGGTAGAGGTTGGGGTGGTGGCTCTCAAGGAGC 1148
Db 61878 AAGTCGCCCTTTAGGTTGCTCGGAGGTAGAGGTTGGGGTGGTGGCTCTCAAGGAGC 61937
Qy 1149 GACTGTCAGATCGCTAGTATGTTCTGTGAACAAATAAAATGATTTACTGTC 1204
Db 61938 GACTGTCAGATCGCTAGTATGTTCTGTGAACAAATAAAATGATTTACTGTC 61993

RESULT 5
LOCUS AP000512 200000 bp DNA PRI 30-MAR-2000
DEFINITION Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
section 11/20.
ACCESSION AP000512
VERSION AP000512.1 GI:5926699
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Shina,S., Tamiya,G., Oka,A. and Inoko,H.
Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region
Published Only in DataBase (1999) in press
2 (bases 1 to 200000)
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
Direct Submission
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases, Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is conducted by Tokai University as a JST sequencing
Team.
COMMENT
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive web site (http://www-alis.tokyo.jst.go.jp/HGS/top.
html) or send email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES
Location/Qualifiers
1. 200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="*6p21.3"
```

```
STS 101440..101648
/note="s1564264:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4566033"
complement(114197..114328)
/standard_name="D6S2115"
/note="SHGC-13451:The location is between each flanking
site of PCR primers."
/db_xref="GDB:738998"
complement(114300..114430)
/note="WI-18721:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4574312"
117236..117362
/note="SGC34396:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4580073"
complement(117712..117836)
/standard_name="D6S2041"
/note="SHGC-10364:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735115"
complement(117961..118093)
/note="SGC32809:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4584619"
142184..142339
/note="K1AA0170:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4586652"
complement(142215..142343)
/note="WI-15445:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4581086"
complement(153977..154114)
/note="SGC34551:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4582352"
complement(165191..165295)
/standard_name="D6S1376"
/note="downstream/upstream:The location is between each
flanking site of PCR primers."
/db_xref="GDB:580497"
188925..188989
/note="pml216:The location is between each flanking site
of PCR primers."
/db_xref="GDB:594325"
complement(197650..197804)
/note="Cda01e04:The location is between each flanking site
of PCR primers."
/db_xref="GDB:443319"
BASE COUNT 52458 a 48336 c 49153 g 50053 t
ORIGIN

Query Match 88.0%; Score 1080.4; DB 51;
Best Local Similarity 91.4%; Pred. No. 9.6e-239;
Matches 1203; Conservative 0; Mismatches 1; Indels 112; Gaps 1;

Qy 1 ATGTGTCACTCTCGAGCTGCCACCGACCATGACCATCTCGAGGCTCGGCTCTGAGATCTTTCACCTTCGACCT 120
Db 97663 ATGTGTCACTCTCGAGCTGCCACCGACCATGACCATCTCGAGGCTCGGCTCGGCTCTGAGATCTTTCACCTTCGACCT 97782
Qy 61 CCTCCACCATCTCCGGGACCCCGCGGGCTCGGCTCTGAGATCTTTCACCTTCGACCT 120
Db 97723 CCTCCACCATCTCCGGGACCCCGCGGGCTCGGCTCTGAGATCTTTCACCTTCGACCT 97782
Qy 121 CTCCGGAGCCCGACGAGGCTCTGCGGGGCGCCCGACGCGCTCTGCGGGGACCGCAAG 180-
Db 97783 CTCCGGAGCCCGACGAGGCTCTGCGGGGCGCCCGACGCGCTCTGCGGGGACCGCAAG 97842
Qy 181 CGCAGCGCAGGCTCTCTACCTCGAGT----- 209
|||||
```

```
Db 15261 TATCCCAACTGGGACTTCGGAGGCAACTTGAACCTCAGAACACTACAGCGGAGACGCCAC 15320
QY 609 CCGGTGCTTACGCGCGGACCGGCGCACAGACACCGCGCATAGAGACCCGAGGCACA 668
Db 15321 CCGGTGCTTACGCGCGGACCGGCGCACAGACACCGCGCATAGAGACCCGAGGCACA 15380
QY 669 GCCCAGCTGGGCTAGGCGCGGTGGGAAGAGAGCGTTCGTTAATTTATTTCTTATTTGCTC 728
Db 15381 GCCCAGCTGGGCTAGGCGCGGTGGGAAGAGAGCGTTCGTTAATTTATTTCTTATTTGCTC 15440
QY 729 CTAATTAATATTTATATGATTTATGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 788
Db 15441 CTAATTAATATTTATATGATTTATGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 15500
QY 789 TATTTATTTTAACTTATGCAAGGCTGTGAGATCTTCCCTCTCTCTCTCTCTCTCTCTCTCT 848
Db 15501 TATTTATTTTAACTTATGCAAGGCTGTGAGATCTTCCCTCTCTCTCTCTCTCTCTCTCTCT 15560
QY 849 TGGTATTTATTTAGCTTTTGGGACTTGGTGGAGAGGACACCTCGGAACCTGCGGCAAAAGT 908
Db 15561 TGGTATTTATTTAGCTTTTGGGACTTGGTGGAGAGGACACCTCGGAACCTGCGGCAAAAGT 15620
QY 909 AGAGAAGAATGGGAGGACTCGGTGGGGGAGGAGCTCCCGCTGGGATGAAGTCTGG 968
Db 15621 AGAGAAGAATGGGAGGACTCGGTGGGGGAGGAGCTCCCGCTGGGATGAAGTCTGG 15680
QY 969 TGGTGGCTCTAAGTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCGCTCTGTCTA 1028
Db 15681 TGGTGGCTCTAAGTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCGCTCTGTCTA 15740
QY 1029 CTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGCTGAGATCTTCCATCTTCTTG 1088
Db 15741 CTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGCTGAGATCTTCCATCTTCTTG 15800
QY 1089 AAGTCGCTTTAGGTGGCTTCGAGGTAGAGGTTTGGGGTTGGTGGCTGTCCAGGAGC 1148
Db 15801 AAGTCGCTTTAGGTGGCTTCGAGGTAGAGGTTTGGGGTTGGTGGCTGTCCAGGAGC 15860
QY 1149 GACTGTGAGATCCGCTAGTATGTTCTGTCAACACAAATAAAATTTACTTCTC 1204
Db 15861 GACTGTGAGATCCGCTAGTATGTTCTGTCAACACAAATAAAATTTACTTCTC 15916

RESULT 4
AB023051
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
DEFINITION clone:876L4, complete sequence.
ACCESSION AB023051
VERSION AB023051.1 GI:5672606
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:876L4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Shihina,T., Tamiya,C., Oka,A., Takishima,N., Yamagata,T.,
Kikawa,E., Iwata,K., Tomizawa,M., Okueki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Bahram,S. and Inoko,H.
Molecular dynamics of MHC genes is unraveled by sequence analysis of
the 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
20027539
TITLE 2 (bases 1 to 90244)
JOURNAL Shihina,T. and Takishima,N.
MEDLINE Direct Submission
AUTHORS Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
TITLE Shihina, Tokai University School of Medicine, Department of
JOURNAL Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail: tshihina@is.icc.u-tokai.ac.jp, Tel: 81-463-93-1121,
Fax: 81-463-94-8884)
```

```
FEATURES             Location/Qualifiers
     source            1..90244
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="876L4"
                     /map="6p21.3"
BASE COUNT   24562 a 21396 c 21572 g 22714 t
ORIGIN
Query Match      88.0%; Score 1080.4; DB 8; Length 90244;
Best Local Similarity 91.4%; Pred. No. 1e-238;
Matches 1203; Conservative 0; Mismatches 1; Indels 112; Gaps 1;
QY 1 ATGTGTCACTCTCCAGCTGCGCACCGGACCATCATCTGAGCGCCCGGACCCCGGCC 60
Db 60678 ATGTGTCACTCTCGAGCTGCCACCGGACCATCATCTGAGCGCCCGGACCCCGGCC 60737
QY 61 CCCTCCACCATCCCGGGACCCCGGGGCTCCGGTCCCTGAGATCTTCACCTTCGACCT 120
Db 60738 CCCTCCACCATCCCGGGACCCCGGGGCTCCGGTCCCTGAGATCTTCACCTTCGACCT 60797
QY 121 CTCGCGAGCCCGACGCGGCCCTGCGGGCGCCCGACGCGCTCTCCGGGCGACCGAAAG 180
Db 60798 CTCGCGAGCCCGACGCGGCCCTGCGGGCGCCCGACGCGCTCTCCGGGCGACCGAAAG 60857
QY 181 CGCAGCGCAGGCTTCTACCTCCGAGT----- 209
Db 60858 CGCAGCGCAGGCTTCTACCTCCGAGTGGTGGTATCCCGAAGTGGCATTCGCGGG 60917
QY 210 ----- 209
Db 60918 GTCCGCTGCCCTGGAGTCACTGGGGAACGACCCGACTCCAGAGCGCTCGACCTGT 60977
QY 210 -----GGTCCGGCGCCAGCTGCCAGTGCAGGAACCGAACCCAGC 248
Db 60978 CTCCTGTTTGTCTCCCTTAGTTCGGCGCCAGCTGCCAGTGCAGGAACCGAACCCAGC 61037
QY 249 CAAAGGCTCTCTTCTGCTGCTCACCATCGTCTTCTGCCACATCTGTATGCTCTGAAGA 308
Db 61038 CAAAGGCTCTCTTCTGCTGCTCACCATCGTCTTCTGCCAGATCTGTATGCTCTGAAGA 61097
QY 309 GGGTGTGCGCGCGCCCTGCTCCAGAGGACGCGCCCTAACGCGCGCATCTCCGCGGCCAC 368
Db 61098 GGGTGTGCGCGCGCCCTGCTCCAGAGGACGCGCCCTAACGCGCGCATCTCCGCGGCCAC 61157
QY 369 CCCTGTGTCCCGCTGCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCT 428
Db 61158 CCCTGTGTCCCGCTGCTCGAGCCCTTTAATCTGACTTCGGAGCCCTGCGACTACGCTCT 61217
QY 429 GGACCTCAGCATTCTCTCCAGCAACACCGCGCGCTTCTAATCTGTACTCTCCCGCACT 488
Db 61218 GGACCTCAGCATTCTCTCCAGCAACACCGCGCGCTTCTAATCTGTACTCTCCCGCACT 61277
QY 489 CCCCCAAAAGATTCGAAAAACCAAAAGAACACCAAGGCGTACCTGGTGGCGGAGAGCG 548
Db 61278 CCCCCAAAAGATTCGAAAAACCAAAAGAACACCAAGGCGTACCTGGTGGCGGAGAGCG 61337
QY 549 TATCCCAACTGGGACTTCGAGGCAACTTGAACCTAGAACACTACAGCGGAGACGCCAC 608
Db 61338 TATCCCAACTGGGACTTCGAGGCAACTTGAACCTAGAACACTACAGCGGAGACGCCAC 61397
QY 609 CCGGTGCTTCCAGCGGGACCGCGCACAGACCGCGGCGCATAGAGACCCGAGGCACA 668
Db 61398 CCGGTGCTTCCAGCGGGACCGCGCACAGACCGCGGCGCATAGAGACCCGAGGCACA 61457
QY 669 GCCCAGCTGGGCTAGGCGCGGTGGGAAGAGAGCGCTCGTTAATTTATTTCTTATTTGCTC 728
Db 61458 GCCCAGCTGGGCTAGGCGCGGTGGGAAGAGAGCGCTCGTTAATTTATTTCTTATTTGCTC 61517
QY 729 CTAATTAATATTTATGATTTATGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 788
```


repeat_region 3941..4070
/rpt_family="Alu"
complement(4449..4724)
/rpt_family="Alu"
repeat_region complement(5117..5389)
/rpt_family="Alu"
repeat_region 7668..7913
/rpt_family="Alu"
repeat_region complement(7924..8226)
/rpt_family="Alu"
repeat_region 8250..8551
/rpt_family="Alu"
repeat_region complement(8893..8973)
/rpt_family="MIR"
repeat_region complement(11829..12096)
/rpt_family="Alu"
repeat_region 12416..12740
/rpt_family="Alu"
repeat_region complement(13181..13247)
/rpt_family="Alu"
repeat_region complement(13498..13647)
/rpt_family="LI"
repeat_region 13869..14279
/rpt_family="Alu"
repeat_region 18033..18284
/rpt_family="Alu"
repeat_region 19178..19451
/rpt_family="Alu"
repeat_region 19458..19749
/rpt_family="Alu"
repeat_region 20862..21002
/rpt_family="MERS"
repeat_region 21566..21832
/rpt_family="Alu"
repeat_region 22231..22521
/rpt_family="Alu"
repeat_region complement(22616..22679)
/rpt_family="Alu"
repeat_region complement(22869..23145)
/rpt_family="Alu"
repeat_region complement(23478..23772)
/rpt_family="Alu"
repeat_region 24184..24372
/rpt_family="Alu"
repeat_region 24529..24813
/rpt_family="Alu"
repeat_region complement(25060..25342)
/rpt_family="Alu"
repeat_region complement(25367..25653)
/rpt_family="Alu"
repeat_region 25787..26076
/rpt_family="Alu"
repeat_region 26491..26762
/rpt_family="Alu"
repeat_region 26866..27143
/rpt_family="Alu"
repeat_region complement(27620..27919)
/rpt_family="Alu"
repeat_region 29525..29719
/rpt_family="MIR"
repeat_region 29762..30333
/rpt_family="Alu"
repeat_region 30365..30653
/rpt_family="Alu"
repeat_region 31430..31722
/rpt_family="Alu"
repeat_region 31765..31857
/rpt_family="Alu"
repeat_region 32265..32563
/rpt_family="Alu"
repeat_region complement(32710..32858)
/rpt_family="Alu"
repeat_region complement(33374..33584)

repeat_region /rpt_family="MER2"
complement(34547..34595)
/rpt_family="MIR"
repeat_region complement(36673..36954)
/rpt_family="Alu"
repeat_region complement(37002..37230)
/rpt_family="Alu"
repeat_region complement(40018..40170)
/rpt_family="Alu"
variation 40452
/note="high quality variation versus 3' overlapping clone"
repeat_region 40509..41109
/rpt_family="Alu"
repeat_region 42609..42892
/rpt_family="Alu"
repeat_region 43240..43370
/rpt_family="MIR"
repeat_region 43493..43779
/rpt_family="Alu"
BASE COUNT 11505 a 10560 c 11226 g 10827 t
ORIGIN
Query Match 88.0%; Score 1080.4; DB 9; Length 44118;
Best Local Similarity 91.4%; Pred No. 1e-238;
Matches 1203; Conservative 0; Mismatches 1; Indels 112; Gaps 1;
Qy 1 ATGTGTACTCTCGCAGCTGCCACCCGACCATGACCATCTCTCGAGGCCCGGACCCGGCC 60
Db 14601 ATGTGTACTCTCGCAGCTGCCACCCGACCATGACCATCTCTCGAGGCCCGGACCCGGCC 14660
Qy 61 CCCTCCACCATCCCGGACCCCGGGGGCTCCGGTCTCTGAGATCTTACCTTCACCTT 120
Db 14661 CCCTCCACCATCCCGGACCCCGGGGGCTCCGGTCTCTGAGATCTTACCTTCACCTT 14720
Qy 121 CTCGGGAGCCGCGACGCGGCCCTCTCGGGGGCCCGACGCTCTCGGGGACCGAAG 180
Db 14721 CTCGGGAGCCGCGACGCGGCCCTCTCGGGGGCCCGACGCTCTCGGGGACCGAAG 14780
Qy 181 CGCAGCCGAGGGTTCCTACCTCTCGAGT----- 209
Db 14781 CGCAGCCGAGGGTTCCTACCTCTCGAGTGTGTAGTATCGCGGAAGTGGGCATTCGCGGG 14840
Qy 210 ----- 209
Db 14841 GTGGCTCCCTGGAGTCACTGGGAACGACCCGACTCCAGAGGCTCGACCTGACCTGT 14900
Qy 210 -----GGTCCGGCGCCAGCTGCCAGTGCAGTGCAGGAGAACCGAACCCACCC 248
Db 14901 CTCCTGTCTTCTCTCCCTTAGGTCCGGCGCCAGCTGCCAGTGCAGGAGAACCGAACCCACCC 14960
Qy 249 CAAAGGCTTCTTCTGTCTGCTCACCATCGTCTTCTGCCAGATCTCTGATGGCTGAAGA 308
Db 14961 CAAAGGCTTCTTCTGTCTGCTCACCATCGTCTTCTGCCAGATCTCTGATGGCTGAAGA 15020
Qy 309 GGGTGTGCGGGGGCCCTTCCTCCAGAGAGACCCCTTACGCCGGATCTCTGGCGCCAC 368
Db 15021 GGGTGTGCGGGGGCCCTTCCTCCAGAGAGAGCCCTTACGCCGGATCTCTGGCGCCAC 15080
Qy 369 CCCTGTGTCCCGGCTCTCGAGCCCTTTAACTCTGAGTCTCGGAGCCCTCGGACTACCGCTCT 428
Db 15081 CCCTGTGTCCCGGCTCTCGAGCCCTTTAACTCTGAGTCTCGGAGCCCTCGGACTACCGCTCT 15140
Qy 429 GGACCTCAGCATTTCTCTCCAGCAACACCGGCCCTTCTTAACCTGTGACTCTCCCGCACT 488
Db 15141 GGACCTCAGCATTTCTCTCCAGCAACACCGGCCCTTCTTAACCTGTGACTCTCCCGCACT 15200
Qy 489 CCCCAGAAAGATCCGAGAAACACCAAGAACACACCGGCTTACCTGTGCGCGAGAGCG 548
Db 15201 CCCCAGAAAGATCCGAGAAACACCAAGAACACACCGGCTTACCTGTGCGCGAGAGCG 15260
Qy 549 TATCCCCAAGTGGGACTTCCGAGGCACTTGAACCTCAGAACACTACAGCGGAGAGCCAC 608
|||||

|||||
619 GAGCCACCGGCTGCTTGGCGGACCGGAGCGCACAGACCGGAGCGCATAGAGACC 678
661 GAGGCACACCCAGCTGGGGCTAGGCCCGGTGGGAGGAGAGCGTGAATTTATTTCT 720
679 GA-GCAGACCCAGCT-GGGCTAGGCCCGGTGGGAGGAGAGCGTGAATTTATTTCT 736
721 TATTGCTCCFAATTAATATTTATATCTATTTATGTCACGTCCTAGGTGATGGAGATCT 780
737 TATTGCTCCFAATTAATATTTATATCTATTTATGTCACGTCCTAGGTGAT-GAGATGT 795
781 GTACGTAATATTTATTTAACTTATGCAAGGGTGTGAGATGTCCTCTGCTGAATGC 840
796 GTACCTAATATTTATTTAACTTATGCAAGGGTGTGAGATGTCCTCTGCTGAATGC 855
841 AGTCTCTGCTGCTATTTATTTAGCTTTGTGGACTGTGCAAGCAGACACCTGGAATCC 900
856 ACCTCTCTGCTGCTATTTATTTAGCTTTGTGGACTGTGCAAGCAGACACCTGGAATCC 915
901 GCGAAAGTAGGCAAGAAATGGGAGGACTTCGGGTGGGAGGAGACCTCCCGGTGGATC 960
916 GCGAAGTAGGCAAGAAATGGGAGGACTTCGGGTGGGAGGAGACCTCCCGGTGGATC 975
961 AACTCTGGTGGTGGTGGTAACTTTAGGAGGTGACATGCTCCAGCA-TCCTCAACTCC 1019
976 AAGTCTGGTGGTGGTGGTAACTTTAGGAGGTGACATGCTCCAGCA-TCCTCAACTCC 1035
1020 GTCTGTACTGTGTGAGACTTCGGCGGACCATTAGAATGAGATCCGTGAGATCCTTC 1079
1036 GTCTGTACTGTGTGAGACTTCGGCGGACCATTAGAATGAGATCCGTGAGATCCTTC 1095
1080 ATCTTCTGAAGTCGCTTTAGGTCGCTCGAGGTAGAGGTGGGGTGGTGGGCTG 1139
1096 ATCTTCTGAAGTCGCTTTAGGTCGCTCGAGGTAGAGGTGGGGTGGTGGGCTG 1155
1140 TCAGGAGGACATCTCGAGATCCGCTAGTATGTTCTGTGAACAAATAAATGATTTA 1199
1156 TCAGGAGGACATCTCGAGATCCGCTAGTATGTTCTGTGAACAAATAAATGATTTA 1215
QY 1200 CTGTC 1204
Db 1216 CTGTC 1220

RESULT 3
AC006165 44118 bp DNA PRI 08-DEC-1998
LOCUS Homo sapiens clone UWGC:y54c125 from 6p21, complete sequence.
DEFINITION AC006165
ACCESSION AC006165
VERSION AC006165.1 GI:3980464
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 44118)
AUTHORS Janer, M., Guillaudoux, T., Vu, Q., Kutayavin, T., Harter, H. and Geraghty, D.E.
TITLE Large scale sequence analysis of the human MHC class I region
JOURNAL Unpublished (1998)
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
2 (bases 1 to 44118)
Geraghty, D.E. and Olson, M.V.
Direct Submission
TITLE Submitted (08-DEC-1998) Human Genome Center, University of
JOURNAL Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145, Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhcrc.org)
Overlapping Sequences:

5': UWGC:y54c283
3': UWGC:y54c222 (Genbank Accession: AC006049)

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 58.8%
DS or two chemistry coverage: 99.5%
Single stranded regions: 2

Sequence Validation:

This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

	BglII		HindIII		NsiI	
	Map	Seq	Map	Seq	Map	Seq
-----	-----	-----	-----	-----	-----	-----
6714.50	6452.00	1823.23	1794.00	5581.12	5401.00	-----
-----	-----	-----	-----	-----	-----	-----
1089.04	1085.00	1246.04	1228.00	10570.24	10123.00	-----
-----	-----	-----	-----	-----	-----	-----
1536.15	1528.00	5570.33	5372.00	10069.29	9626.00	-----
-----	-----	-----	-----	-----	-----	-----
3623.08	3537.00	1670.92	1630.00	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
3784.08	3621.00	1270.76	1280.00	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
6834.00	6689.00	4235.79	4109.00	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
11234.36	10874.00	1171.00	1151.00	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
762.43	762.00	3226.11	3096.00	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
2727.38	2653.00	10501.64	10080.00	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
1691.42	1684.00	4208.44	4052.00	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
1295.60	1277.00	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
746.87	734.00	-----	-----	-----	-----	-----

FEATURES

Source	Location/Qualifiers
1. 44118	/organism="Homo sapiens"
/db_xref="taxon:9606"	
/chromosome="6"	
/map="6p21"	
/clone="CGM1:D100B8"	
/sub_clone="UWGC:y54c125"	
/cell_line="CGM1"	
/clone_lib="Wash U YAC library"	
218..405	/rpt_family="MER3"
400..999	/rpt_family="Alu"
1797..2378	/rpt_family="Alu"
2649..2956	/rpt_family="Alu"
complement(3372..3894)	/rpt_family="Alu"
repeat_region	
repeat_region	
repeat_region	
repeat_region	
repeat_region	

c 22 64 5.2 209887 71 AC079424 Mus muscu
23 63.4 5.2 64739 62 AC022717 Homo sapi
c 24 62.6 5.1 300695 71 AC079431 Mus muscu
c 25 60.8 5.0 187738 66 AC027682 Homo sapi
26 59.6 4.9 134940 60 AC018939 Homo sapi
27 59.4 4.8 50602 56 AC011100 Homo sapi
c 28 59.4 4.8 120379 56 AC011429 Homo sapi
c 29 59.4 4.8 220469 70 AC074307 Mus muscu
c 30 59.2 4.8 62649 62 AC022552 Homo sapi
31 58.8 4.8 78765 61 AC021584 Homo sapi
32 58.6 4.8 259451 58 AC015605 Homo sapi
33 58.4 4.8 149915 61 AC021637 Homo sapi
34 58.2 4.7 75628 61 AC021793 Homo sapi
c 35 58.2 4.7 135217 71 AC079081 Homo sapi
36 58 4.7 79023 61 AC021951 Homo sapi
37 58 4.7 87440 61 AC021226 Homo sapi
38 58 4.7 150695 63 AC023576 Homo sapi
c 39 57.8 4.7 200580 72 AC084064 Homo sapi
40 57.6 4.7 141892 62 AC023197 Homo sapi
41 57.4 4.7 76795 63 AC023515 Homo sapi
42 57.2 4.7 840 7 CNS01BNS Botrytis
43 57.2 4.7 73948 62 AC022556 Homo sapi
c 44 57.2 4.7 258174 71 AC079429 Mus muscu
c 45 57 4.6 249262 71 AC079430 Mus muscu

ALIGNMENTS

RESULT 1
HSDIF2 HSDIF2 1230 bp mRNA PRI 12-AUG-1997
LOCUS Homo sapiens mRNA for DIF-2 protein.
DEFINITION Y14551
ACCESSION Y14551.1 GI:2660550
VERSION dif-2 gene.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1230)
AUTHORS Pietzsch,A., Buchler,C., Aslanidis,C. and Schmitz,G.
TITLE Identification and characterization of a novel monocyte/macrophage
differentiation-dependent gene that is responsive to
lipopolysaccharide, ceramide, and lysophosphatidylcholine
JOURNAL Biochem. Biophys. Res. Commun. 235 (1), 4-9 (1997)
MEDLINE 97339426
REFERENCE 2 (bases 1 to 1230)
AUTHORS Pietzsch,A.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1997) A. Pietzsch, Institute for Clinical
Chemistry and Laboratory Medicine, University of Regensburg,
Franz-Josef-Strauss-Allee 11, 93053 Regensburg, FRG
REFERENCE 3 (bases 1 to 1230)
AUTHORS Kondratyev,A.D., Chung,K.N. and Jung,M.O.
TITLE Identification and characterization of a radiation-inducible
glycosylated human early-response gene
JOURNAL Cancer Res. 56 (7), 1498-1502 (1996)
MEDLINE 96181295
FEATURES Location/Qualifiers
source 1..1230
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="leukocytes"
1..11
/gene="dif-2"
1..1230
/gene="dif-2"
12..482
/gene="dif-2"
/codon_start=1
/product="DIF-2 protein"
/protein_id="CAA74886.1"

/db_xref="GI:2660551"
/db_xref="SWISS-PROT:P46695"
/translation="MCHSRSCHPMTILOAPTPAPSTIPGPRRCSPGPEITFDPLPEP
AAPAGRPASRGRKSRRLVRYRQQLPVEPNPARKRLLELLLTIVFCOTIMAE
EGVPAPLPEDAPNAASLAPTPVSPVLEPFLTSEPSYALDLSLTFLOOHPAAP"
172
conflict /gene="dif-2"
/citation=[3]
/replace="g"
328
conflict /gene="dif-2"
/citation=[3]
/replace="g"
483..>1230
/gene="dif-2"
673..675
conflict /gene="dif-2"
/citation=[3]
/replace="ag"
687..689
conflict /gene="dif-2"
/citation=[3]
/replace="tg"
783..785
conflict /gene="dif-2"
/citation=[3]
/replace="tg"
1020..1021
conflict /gene="dif-2"
/citation=[3]
/replace="att"
1120
conflict /gene="dif-2"
/citation=[3]
/replace="g"
1133
conflict /gene="dif-2"
/citation=[3]
/replace="t"
BASE COUNT 259 a 367 c 342 g 262 t
ORIGIN
Query Match 98.4%; Score 1207.8; DB 53; Length 1230;
Best Local Similarity 99.4%; Pred. No. 5.7e-268;
Matches 1212; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1 ATGTGTACTCTCGAGCTGCCACCGACCATGACATCTCGAGCCCGCCGCGCC 60
Db 12 ATGTGTACTCTCGAGCTGCCACCGACCATGACATCTCGAGCCCGCCGCGCC 71
Qy 61 CCTCCACCATCCCGGACCCCGGGGCTCCGGTCTCGATCTTACCTTCGACCT 120
Db 72 CCTCCACCATCCCGGACCCCGGGGCTCCGGTCTCGATCTTACCTTCGACCT 131
Qy 121 CTCGGGAGCCCGGACCCCGCCCTGCGGGGCGCCCGACGCGCTCTCGGGGACCGAAG 180
Db 132 CTCGGGAGCCCGGACCCCGCCCTGCGGGGCGCCCGACGCGCTCTCGGGGACCGAAG 191
Qy 181 CGCAGCCGAGGGTCTCTACCTCGAGTGTGTCGGGCGCCAGCTGCCAGTCGAGANCCG 240
Db 192 CGCAGCCGAGGGTCTCTACCTCGAGTGTGTCGGGCGCCAGCTGCCAGTCGAGANCCG 251
Qy 241 AACCCAGCAAAAGGCTTCTCTTCTGCTCACCATCTTCTGTCGAGATCTTCTGATG 300
Db 252 AACCCAGCAAAAGGCTTCTCTTCTGCTCACCATCTTCTGTCGAGATCTTCTGATG 311
Qy 301 GCTGAAGAGGGTGTGCGCGCGCCCTGCTCCAGAGGAGCGCCCTAACGCCCATCCCTG 360
Db 312 GCTGAAGAGGGTGTGCGCGCGCCCTGCTCCAGAGGAGCGCCCTAACGCCCATCCCTG 371
Qy 361 GCGCCACCCCTGTGTCTCCCGCCCTGCTCGAGCCCTTAACTGACTTCGAGGCCCTCGGAC 420

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 12:11:54 ; Search time 1900.13 Seconds
(without alignments)
3307.456 Million cell updates/sec

Title: US-08-799-910-9
Perfect score: 1228
Sequence: 1 ATGCTCACTCTCGACGCTG.....AAAAAAAAAAAAAACTCGAG 1228

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues 2236266
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_ph.*
- 6: gb_pl1.*
- 7: gb_pl2.*
- 8: gb_pr1.*
- 9: gb_pr2.*
- 10: gb_pr3.*
- 11: gb_ro.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: em_fun.*
- 15: em_hum1.*
- 16: em_hum2.*
- 17: em_in.*
- 18: em_om.*
- 19: em_or.*
- 20: em_ov.*
- 21: em_pat.*
- 22: em_ph.*
- 23: em_pi.*
- 24: em_ro.*
- 25: em_sts.*
- 26: em_sy.*
- 27: em_un.*
- 28: em_vi.*
- 29: gb_ba3.*
- 30: gb_in1.*
- 31: gb_in2.*
- 32: gb_in3.*
- 33: gb_pl3.*
- 34: gb_pr4.*
- 35: em_ba1.*
- 36: em_ba2.*
- 37: em_htg1.*
- 38: em_htg2.*
- 39: em_htg3.*
- 40: em_htg4.*
- 41: em_htg5.*
- 42: em_htg6.*
- 43: em_htg7.*

- 44: em_htg8.*
- 45: em_htg9.*
- 46: em_htg10.*
- 47: em_hum3.*
- 48: em_hum4.*
- 49: em_hum5.*
- 50: em_hum6.*
- 51: gb_pr5.*
- 52: gb_pr6.*
- 53: gb_pr7.*
- 54: gb_htg1.*
- 55: gb_htg2.*
- 56: gb_htg3.*
- 57: gb_htg4.*
- 58: gb_htg5.*
- 59: gb_htg6.*
- 60: gb_htg7.*
- 61: gb_htg8.*
- 62: gb_htg9.*
- 63: gb_htg10.*
- 64: gb_htg11.*
- 65: gb_htg12.*
- 66: gb_htg13.*
- 67: gb_htg14.*
- 68: gb_htg15.*
- 69: gb_htg16.*
- 70: gb_htg17.*
- 71: gb_htg18.*
- 72: gb_htg19.*
- 73: gb_htg20.*
- 74: gb_htg21.*
- 75: gb_htg22.*
- 76: gb_htg23.*
- 77: gb_sts1.*
- 78: gb_sts2.*
- 79: gb_vil.*
- 80: gb_vil2.*
- 81: gb_pat1.*
- 82: gb_pat2.*
- 83: em_htg0.*
- 84: gb_htg24.*
- 85: gb_pr8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match		Length	DB	ID	Description
		Match	Length				
1	1207.8	98.4	1230	53	HSDF2		Y14551 Homo sapien
2	1152.2	93.8	1223	85	S81914		S81914 IEX-1-radia
3	1080.4	88.0	44118	9	AC006165		AC006165 Homo sapi
4	1080.4	88.0	90244	8	AB023051		AB023051 Homo sapi
5	1080.4	88.0	200000	51	AP000512		AP000512 Homo sapi
6	1048.6	85.4	1309	10	AF039067		AF039067 Homo sapi
7	1048.6	85.4	1693	10	AF071596		AF071596 Homo sapi
8	979.8	79.8	1864	53	HSPRG1		X96438 H.sapiens p
9	471	38.4	477	10	AF083421		AF083421 Homo sapi
10	298.8	24.3	343	52	HSAD227914		AJ227914 Homo sapi
11	242.4	19.7	1758	11	RNPRG1		X96437 R.norvegicu
12	210.6	17.1	203476	70	AC074150		AC074150 Mus muscu
13	209	17.0	1938	11	MMGLY96		X67644 M.musculus
14	202	16.4	187868	62	AC022301		AC022301 Mus muscu
15	151	12.3	297	81	A75448		A75448 Sequence 11
16	151	12.3	297	81	A78427		A78427 Sequence 11
17	122	9.9	122	78	G62116		G62116 1899 Human
18	74	6.0	205512	74	AL355388		AL355388 Homo sapi
19	70.4	5.7	267	81	A75468		A75468 Sequence 11
20	70.4	5.7	267	81	A78447		A78447 Sequence 11
21	64	5.2	83276	65	AC026631		AC026631 Homo sapi

UUUUUUUU

DR
DR
DR
DR
DR
KW
KW
FT
FT
FT
FT
FT
SQ

Qy
Db

Sea.
Job

RA Angrand P.O., Valvatne H., Jeannougin F., Adamson A.,
RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,
RA Lamerdin J., Chambon P., Losson R., Stewart A., Aasland R.,
RT "Mammalian trithorax" and ASH1-like proteins: putative chromatin
RT regulators which contain PHD fingers and SET domains.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A. (LONG ISOFORM).
RP Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,
RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Garges J., Danganan L.,
RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
RA Carrano A.V.;
RT "Sequence analysis of a 1 Mb region in human 19q13.1";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 816-2715 FROM "N.A. (LONG ISOFORM).
RP TISSUE-BRAIN;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RN [4]
RN SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).
RP TISSUE-TESTIS, AND LEUKOCYTE;
RX MEDLINE=20105772; PubMed=10637508;
RA Huntsman D.G., Chin S.-F., Muleris M., Batley S.J., Collins V.P.,
RA Wiedemann L.M., Aparicio S., Caldas C.;
RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps
RT to 19q13.1 and is amplified in solid tumor cell lines.";
RL Oncogene 18:7975-7984(1999).
RN [5]
RN PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).
RP TISSUE-PLACENTA, AND BONE MARROW;
RX MEDLINE=99339983; PubMed=10409430;
RA Fitzgerald K.T., Diaz M.O.;
RT "MLL2: A new mammalian member of the trx/MLL family of genes.";
RL Genomics 59:187-192(1999).
CC -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.
CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,
CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL
CC BLOOD LYMPHOCYTES, AND PLACENTA.
CC -!- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.
CC -!- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 PHD-FINGER DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ007041; CAB45385.1; -;
DR EMBL: AD000671; -; NOT_ANNOTATED_CDS.
DR EMBL: AB002302; BAA20763.2; -;
DR EMBL: AF186605; AAD56420.1; -;
DR EMBL: AF104918; AAD17932.1; -;
DR EMBL: AF105279; AAD26113.1; -;
DR EMBL: AF105280; AAD26112.1; -;
DR INTERPRO: IPR001214; -;
DR INTERPRO: IPR001965; -;
DR INTERPRO: IPR002857; -;

DR PFAM: PF00628; PHD; 3.
DR PFAM: PF00856; SET; 1.
DR PFAM: PF02008; zf-CXXC; 1.
KW DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;
KW Transcription regulation; Alternative splicing.
FT DNA_BIND 37 44 A..T HOOK (BY SIMILARITY).
FT DNA_BIND 110 117 A..T HOOK (BY SIMILARITY).
FT DNA_BIND 1207 365 A..T HOOK (BY SIMILARITY).
FT DOMAIN 1303 1252 PHD 1.
FT DOMAIN 1337 1396 PHD 2.
FT DOMAIN 1449 1471 PHD 3.
FT DOMAIN 2586 2715 BROMODOMAIN (DIVERGENT).
FT DOMAIN 26 37 SET DOMAIN.
FT DOMAIN 248 255 POLY-GLY.
FT DOMAIN 362 398 POLY-PRO.
FT DOMAIN 402 771 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 808 812 PRO-RICH.
FT DOMAIN 1963 1970 POLY-GLN.
FT DOMAIN 2251 2259 POLY-PRO.
FT VARSPPLIC 532 582 VSARSRVKTPRRRNDEDPKPKPKVSPVLRPPITTSPP
FT VPOEPAPVPS -> PLUSQLLPMTLQLSLGQWNAPTTS
FT ACIDSLPLMSPLLLRPRCLPLTGLQL (IN ISOFORM
FT TRUNCATED).
FT VARSPLIC 583 2715 MISSING (IN ISOFORM TRUNCATED).
FT CONFLICT 834 834 K -> E (IN REF. 5).
FT CONFLICT 941 941 S -> Y (IN REF. 5).
FT CONFLICT 1317 1317 E -> Q (IN REF. 5).
FT CONFLICT 1362 1362 H -> Y (IN REF. 5).
FT CONFLICT 1438 1438 D -> N (IN REF. 5).
FT CONFLICT 2622 2622 D -> H (IN REF. 5).
SQ SEQUENCE 2715 AA; 293511 MW; C0615B981BBB7BF CRC64;

Query Match 16.0%; Score 70.5; DB 1; Length 2715;
Best Local Similarity 34.5%; Pred. No. 43;
Matches 19; Conservative 8; Mismatches 27; Indels 1; Gaps 1;

QY 8 EEPNPAKRLLFLLLTLTFVFCOILMAEEGVAPLPEDAPNNAASLAPT-PVSPVLEP 61
Db 2182 EPPKATSKILLVKNKGQVFVKMAGEGEPVPPVKQPLPTTISAPTTSWTLP 2236

RESULT 9

ID_FAS2_YEAST STANDARD; PRT: 1894 AA.
AC P19097;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FATTY ACID SYNTHASE, SUBUNIT ALPHA (EC 2.3.1.86) [INCLUDES:
DE EC 1.1.1.100; EC 2.3.1.41].
GN FAS2 OR YPL231W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88315020; PubMed=2900835;
RA Mohamed A.H., Chirala S.S., Mody N.H., Huang W.Y., Wakil S.J.;
RT "Primary structure of the multifunctional alpha subunit protein of
RT yeast fatty acid synthase derived from FAS2 gene sequence.";
RL J. Biol. Chem. 263:12315-12325(1988).
RN [2]
RN MUTAGENESIS OF GLY-1257.
RP STRAIN=S288C;
RX MEDLINE=94316198; PubMed=8041367;
RA Inokoshi J., Tomoda H., Hashimoto H., Watanabe A., Takeshima H.,
RA Omura S.;
RT "Cerulenin-resistant mutants of Saccharomyces cerevisiae with an
RT altered fatty acid synthase gene.";
RL Mol. Genet. 244:90-96(1994).
CC -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF

Query Match 16.4%; Score 72.5; DB 1; Length 272;
 Best Local Similarity 30.1%; Pred. NO. 2.2;
 Matches 22; Conservative 16; Mismatches 22; Indels 13; Gaps 4;

OY 4 QLPVEENPAKRLFLLLTIVFCO---ILMAEEGVAPLPPEADAPNAASLAPTPVSPVLE 60
 Db 5 KLPLPSVIRILFYLLMLLFCOOLAMIFWRIGL-----PDNAP-VSSVQITPAQARQQ 58
 OY 61 PNLTSPEPSDYAL 73
 Db 59 PVTL-----NDFTL 67

RESULT 6
 GPBA_HUMAN STANDARD; PRT; 626 AA.

AC P07359;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PLATELET GLYCOPROTEIN IB ALPHA CHAIN PRECURSOR (GP-IB ALPHA) (GP1BA)
 DE (CD42B-ALPHA) (CD42B) [CONTAINS: GLYCOCALICIN].
 GN GP1BA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=87289655; PubMed=3303030;
 RA Lopez J.A., Chung D.W., Fujikawa K., Hagen F.S., Papayannopoulou T.,
 RA Roth G.J.;
 RT "Cloning of the alpha chain of human platelet glycoprotein Ib: a
 RT transmembrane protein with homology to leucine-rich alpha 2-
 RT glycoprotein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:5615-5619(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89025874; PubMed=2845978;
 RA Wenger R.H., Kieffer N., Wicki A.N., Clemetson K.J.;
 RT "Structure of the human blood platelet membrane glycoprotein Ib alpha
 RT gene.";
 RL Biochem. Biophys. Res. Commun. 156:389-395(1988).
 RN [3]
 RP SEQUENCE OF 17-315.
 RX MEDLINE=87289654; PubMed=3497398;
 RA Titani K., Takio K., Handa M., Ruggeri Z.M.;
 RT "Amino acid sequence of the von Willebrand factor-binding domain of
 RT platelet membrane glycoprotein Ib.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:5610-5614(1987).
 RN [4]
 RP DISULFIDE BONDS.
 RX MEDLINE=91301149; PubMed=2070794;
 RA Hess D., Schaller J., Rickli E.E., Clemetson K.J.;
 RT "Identification of the disulphide bonds in human platelet
 RT glycoalbumin.";
 RL Eur. J. Biochem. 199:389-393(1991).
 RN [5]
 RP VARIANT SIBA.
 RX MEDLINE=92265982; PubMed=1586750;
 RA Murata M., Furihata K., Ishida F., Russell S.R., Ware J.,
 RA Ruggeri Z.M.;
 RT "Genetic and structural characterization of an amino acid dimorphism
 RT in glycoprotein Ib alpha involved in platelet transfusion
 RT refractoriness.";
 RL Blood 79:3086-3090(1992).
 RN [6]
 RP VARIANT BSS PHE-73.
 RX MEDLINE=92110577; PubMed=1730088;
 RA Miller J.L., Lyle V.A., Cunningham D.;
 RT "Mutation of leucine-57 to phenylalanine in a platelet glycoprotein
 RT Ib alpha leucine tandem repeat occurring in patients with an

RT autosomal dominant variant of Bernard-Soulier disease.";
 RL Blood 79:439-446(1992).
 RN [7]
 RP VARIANT BSS VAL-172.
 RX MEDLINE=9338851; PubMed=7690774;
 RA Ware J., Russell S.R., Marchese P., Murata M., Mazzucato M.,
 RA de Marco L., Ruggeri Z.M.;
 RT "point mutation in a leucine-rich repeat of platelet glycoprotein Ib
 RT alpha resulting in the Bernard-Soulier syndrome.";
 RL J. Clin. Invest. 92:1213-1220(1993).
 RN [8]
 RP VARIANT BSS SER-225.
 RX MEDLINE=95118882; PubMed=7819107;
 RA Simsek S., Norris P., Lozano M., Pico M., von Dem Borne A.E.G.K.,
 RA Ribera A., Gallardo D.;
 RT "Cys209 Ser mutation in the platelet membrane glycoprotein Ib alpha
 RT gene is associated with Bernard-Soulier syndrome.";
 RL Br. J. Haematol. 88:839-844(1994).
 RN [9]
 RP VARIANT PSEUDO-VWD VAL-249.
 RX MEDLINE=91271273; PubMed=2052556;
 RA Miller J.L., Cunningham D., Lyle V.A., Finch C.N.;
 RT "Mutation in the gene encoding the alpha chain of platelet
 RT glycoprotein Ib in platelet-type von Willebrand disease.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4761-4765(1991).
 RN [10]
 RP VARIANT PSEUDO-VWD VAL-249.
 RX MEDLINE=93253059; PubMed=8486780;
 RA Murata M., Russell S.R., Ruggeri Z.M., Ware J.;
 RT "Expression of the phenotypic abnormality of platelet-type von
 RT Willebrand disease in a recombinant glycoprotein Ib alpha fragment.";
 RL J. Clin. Invest. 91:2133-2137(1993).
 RN [11]
 RP VARIANT PSEUDO-VWD VAL-255.
 RX MEDLINE=93214031; PubMed=8384898;
 RA Russell S.D., Roth G.J.;
 RT "pseudovon Willebrand disease: a mutation in the platelet
 RT glycoprotein Ib alpha gene associated with a hyperactive surface
 RT receptor.";
 RL Blood 81:1787-1791(1993).
 RN [12]
 RP VARIANT BSS LEU-195 DEL.
 RX MEDLINE=95178321; PubMed=7873390;
 RA de la Salle C., Baas M.-J., Lanza F., Schwartz A., Hansu D.,
 RA Chevalier J., Gachet C., Briquel M.-E., Cazenave J.-P.;
 RT "A three-base deletion removing a leucine residue in a leucine-rich
 RT repeat of platelet glycoprotein Ib alpha associated with a variant of
 RT Bernard-Soulier syndrome (Nancy I).";
 RL Br. J. Haematol. 89:386-396(1995).
 CC -1- FUNCTION: GP-IB, A SURFACE MEMBRANE PROTEIN OF PLATELETS,
 CC PARTICIPATES IN THE FORMATION OF PLATELET PLUGS BY BINDING TO VON
 CC WILLEBRAND FACTOR, WHICH IS ALREADY BOUND TO THE SUBENDOTHELIUM.
 CC -1- SUBUNIT: GP-IB ALPHA AND BETA ARE DISULFIDE LINKED. GP-IX IS
 CC COMPLEXED WITH THE GP-IB HETERODIMER VIA A NON COVALENT LINKAGE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- PTM: GLYCOCALICIN, WHICH IS APPROXIMATELY COEXTENSIVE WITH THE
 CC EXTRACELLULAR PART OF THE MOLECULE, IS CLEAVED OFF BY CALPAIN
 CC DURING PLATELET LYSIS.
 CC -1- POLYMORPHISM: POSITION 161 IS ASSOCIATED WITH PLATELET-SPECIFIC
 CC ALLOANTIGEN SIBA. SIBA(-) HAS THR-161 AND SINA(+) HAS MET-161.
 CC SIBA IS INVOLVED IN NEONATAL ALLOIMMUNE THROMBOCYTOPENIA (NATP).
 CC -1- DISEASE: DEFECTS IN GP1BA ARE ONE OF THE CAUSES OF BERNARD-SOULIER
 CC SYNDROME (BSS). BSS PATIENTS HAVE UNUSUALLY LARGE PLATELETS AND
 CC HAVE A CLINICAL BLEEDING TENDENCY.
 CC -1- DISEASE: DEFECTS IN GP1BA ARE ONE OF THE CAUSES OF VON WILLEBRAND
 CC DISEASE (VWD) KNOWN AS PLATELET-TYPE VON WILLEBRAND DISEASE OR
 CC PSEUDO-VON WILLEBRAND DISEASE. THIS AUTOSOMAL DOMINANT BLEEDING
 CC DISORDER IS CAUSED BY AN INCREASED AFFINITY OF GP-IB FOR SOLUBLE
 CC VWF RESULTING IN IMPAIRED HEMOSTATIC FUNCTION DUE TO THE REMOVAL
 CC OF VWF FROM THE CIRCULATION.
 CC -1- MISCELLANEOUS: PLATELET ACTIVATION APPARENTLY INVOLVES DISRUPTION
 CC OF THE MACROMOLECULAR COMPLEX OF GP-IB WITH THE PLATELET
 CC GLYCOPROTEIN IX (GP-IX) AND DISSOCIATION OF GP-IB FROM THE

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; S81914; AAR36278.1; -;
DR EMBL; X96438; CAA65304.1; -;
DR EMBL; Y14551; CAA74886.1; -;
DR EMBL; T49531; -; NOT_ANNOTATED_CDS.
DR MIM; 602996; -;
KW Glycoprotein; Transmembrane; Signal-anchor
FT DOMAIN 1 82
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 100 156 * EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 54 54 A -> G (IN REF. 1).
FT CONFLICT 106 106 P -> R (IN REF. 1).
SQ SEQUENCE 156 AA; 16928 MW; 83C06116C80B9240 CRC64;

Query Match 100.0%; Score 442; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.6e-37;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTPVSPVLE 60
DB 71 VRRQLPVEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTPVSPVLE 130

QY 61 PNLTLSEPSDYALDLSTFLQHPAAAF 86
DB 131 PNLTLSEPSDYALDLSTFLQHPAAAF 156

RESULT 2
ID IEX1_MOUSE STANDARD; PRT; 153 AA.
AC P46694;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY
DE PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN).
GN IER3 OR GLY96.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=93173526; PubMed=8437864;
RA Charles C.H., Yoon J.K., Simske J.S., Lau L.F.;
RT "Genomic structure, cDNA sequence, and expression of gly96, a growth
RT factor-inducible immediate-early gene encoding a short-lived
RT glycosylated protein.";
RL Oncogene 8:1797-801(1993).
CC -!- FUNCTION: NOT KNOWN; EXPRESSED DURING THE G0-G1 TRANSITION OF THE
CC CELL CYCLE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE LUNG, TESTES
CC AND THE UTERUS.
CC -!- INDUCTION: BY SERUM GROWTH FACTORS AND TPA.
CC -!- PTM: GLYCOSYLATED.
CC -!- SIMILARITY: STRONG, TO HUMAN ORTHOLOG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X67644; -; NOT_ANNOTATED_CDS.
DR PIR; S33363; S33363.
DR MGD; MGI:104814; IER3.
KW Glycoprotein; Transmembrane; Signal-anchor
FT DOMAIN 1 85
FT SIGNAL-ANCHOR (POTENTIAL).
FT (POTENTIAL).
FT DOMAIN 103 153
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 153 AA; 16875 MW; F4429E3B0120DE11 CRC64;

Query Match 61.8%; Score 273; DB 1; Length 153;
Best Local Similarity 68.2%; Pred. No. 1.8e-20;
Matches 58; Conservative 5; Mismatches 14; Indels 8; Gaps 2;

QY 1 VRRQLPVEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTPVSPVLE 56
DB 71 VRRQLPVEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTPVSPVLE 128

QY 57 --PVLEPNLTSEPSDYALDLSTFL 79
DB 129 APPVLEPLNLTSSESDYALDLKAF 153

RESULT 3
ID SOS_DROME STANDARD; PRT; 1595 AA.
AC P28675;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SON OF SEVENLESS PROTEIN.
GN SOS
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OREGON-R;
RX MEDLINE=92141820; PubMed=1736363;
RA Bonfini L., Karlovich C.A., Dasgupta C., Banerjee U.;
RT "The Son of sevenless gene product: a putative activator of Ras.";
RL Science 255:603-606(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92034991; PubMed=1934068;
RA Simon M.A., Bowtell D.D.L., Dodson G.S., Lavery T.R., Rubin G.M.;
RT "Ras1 and a putative guanine nucleotide exchange factor perform
RT crucial steps in signaling by the sevenless protein tyrosine
RT kinase.";
RL Cell 67:701-716(1991).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. SOS IS
CC IMPLICATED IN NEURONAL DEVELOPMENT.
CC -!- SUBUNIT: MAY FORM A COMPLEX WITH SEVENLESS AND DRK.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M83931; AAB04680.1; -;
DR EMBL; M77501; AAA28904.1; -;
DR PIR; A41216; A41216.
DR HSPSP; P04002; 1WFA.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: January 30, 2001, 23:12:24 ; Search time 43.61 seconds
(without alignments)
63.685 Million cell updates/sec

Title: US-08-799-910-10_COPY_71_156

Perfect score: 442

Sequence: 1 VRRQLPVEEPNPAKRLFL.....EPSDYALDLSTFLQQHPAAF 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	442	100.0	156	1	IEX1_HUMAN
2	273	61.8	153	1	IEX1_MOUSE
3	78.5	17.8	1595	1	SOS-DROME
4	73.5	16.6	272	1	GSPC_ERWCH
5	72.5	16.4	272	1	GSOC_ERWCH
6	71.5	16.2	626	1	GPBA_HUMAN
7	70.5	16.0	605	1	BRL1_EBV
8	70.3	16.0	2715	1	TRX2_HUMAN
9	69.5	15.7	1894	1	FAS2_YEAST
10	69	15.6	1402	1	IF4G_RABIT
11	68.5	15.5	181	1	C21Q_HUMAN
12	68.5	15.5	882	1	CAD1_HUMAN
13	68.5	15.5	1337	1	PTPB_HUMAN
14	68.5	15.5	1844	1	POLR_TYMV
15	68	15.4	1395	1	IF4G_HUMAN
16	67	15.2	283	1	EXTN_SORBI
17	67	15.2	586	1	LREL_YEAST
18	67	15.2	4289	1	TENX_HUMAN
19	66.5	15.0	3149	1	TEGU_EBV
20	66	14.9	621	1	ILVB_MYCAV
21	65.5	14.8	325	1	MODD_MYCBO
22	65.5	14.8	325	1	MODD_MYCTU
23	65.5	14.8	329	1	EXBB_PSEPU
24	65.5	14.8	812	1	EBN3_EBV
25	65.5	14.8	1093	1	AF17_HUMAN
26	65	14.7	209	1	CX11_HUMAN
27	65	14.7	328	1	ZIFA_ECOLI
28	65	14.7	427	1	TEF3_HUMAN
29	65	14.7	724	1	P85B_BOVIN
30	64.5	14.6	686	1	MYBB_CHICK
31	64.5	14.6	1183	1	DRFL_RAT
32	64	14.5	510	1	C933_SOYBN
33	64	14.5	510	1	IA1_HUMAN

34 64 14.5 537 1 PTN5_HUMAN
35 64 14.5 541 1 PTN5_MOUSE
36 64 14.5 590 1 PTZ1_RHIME
37 64 14.5 966 1 SSG6_YEAST
38 63.5 14.4 443 1 HXA3_HUMAN
39 63.5 14.4 540 1 NUSA_MYCPN
40 63.5 14.4 816 1 GAIF_NEUCR
41 63.5 14.4 1776 1 POLR_OYMV
42 63.5 14.4 1844 1 POLR_TYMV
43 63 14.3 323 1 PF27_MOUSE
44 63 14.3 420 1 EF1A_HALMA
45 63 14.3 598 1 NR41_HUMAN

P54829 homo sapien
P54830 mus musculu
P30327 thizobium m
P14922 saccharomyc
O43365 homo sapien
P75591 mycoplasma
P11638 neurospora
P20127 ononis yell
P28477 turnip yell
P52875 mus musculu
P16018 haloarcula
P22736 homo sapien

ALIGNMENTS

RESULT 1
IEX1_HUMAN

ID IEX1_HUMAN STANDARD: PRT: 156 AA.
AC P46695; Q93044; Q92691;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY
DE PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN) (PACAP-RESPONSIVE
DE GENE 1 PROTEIN) (PRG1 PROTEIN) (DIFFERENTIATION-DEPENDENT GENE 2
DE PROTEIN) (DIF-2 PROTEIN).
GN IER3 OR IEX1 OR PRG1 OR DIF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=96181295; PubMed=8603392;
RA Kondratyev A.D., Chung K.-N., Jung M.O.;
RT "Identification and characterization of a radiation-inducible
RT glycosylated human early-response gene.";
RL Cancer Res. 56:1498-1502(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96221139; PubMed=8653710;
RA Schaefer H., Trausold A., Siegel E.G., Folsch U.R., Schmidt W.E.;
RT "PRG1: a novel early-response gene transcriptionally induced by
RT pituitary adenylate cyclase activating polypeptide in a pancreatic
RT carcinoma cell line.";
RL Cancer Res. 56:2641-2641(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97339426; PubMed=9196025;
RA Pietzsch A., Buechler C., Aslanidis C., Schmitz G.;
RT "Identification and characterization of a novel monocyte/macrophage
RT differentiation-dependent gene that is responsive to
RT lipopolysaccharide, ceramide, and lysophosphatidylcholine.";
RL Biochem. Biophys. Res. Commun. 235:4-9(1997).
RN [4]
RP PRELIMINARY SEQUENCE OF 1-106 FROM N.A.
RC TISSUE=PLACENTA;
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M.,
RA Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,
RA Rifkin L., Rohlfing T., Tan F., Trevaskis E., Waterston R.,
RA Williamson A., Wohldmann P., Wilson R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -!- INDUCTION: BY RADIATION, TPA, OKADAIC ACID AND TNF-ALPHA.
CC -!- PTM: GLYCOSYLATED.
CC -!- SIMILARITY: STRONG, TO MOUSE ORTHOLOG.
CC -!- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman C.S., Pan S., Poilard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [5]
RP SEQUENCE OF 281-349 FROM IN A.
RX MEDLINE=93066327; PubMed=1438276;
RA Sommer R.J., Retzlaff M., Goerlich K., Sander K., Tautz D.;
RT "Evolutionary conservation pattern of zinc-finger domains of
RT *Drosophila* segmentation genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10782-10786(1992).
RN [6]
RP POLYMORPHISM.
RA Tautz D., Nigro L.;
RT "Microevolutionary divergence pattern of the segmentation gene
RT hunchback in *Drosophila*.";
RL Mol. Biol. Evol. 15:1403-1411(1998).
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC OF HEAD STRUCTURES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: IN EMBRYO, EXPRESSION OF MATERNAL TRANSCRIPT
CC IS HIGHEST IN ANTERIOR REGION. ZYGOTIC TRANSCRIPT IS EXPRESSED IN
CC ANTERIOR REGION UNTIL THE BEGINNING OF GASTRULATION AND IN
CC POSTERIOR REGION UNTIL EARLY GASTRULATION. AFTER THIS, IT IS
CC EXPRESSED IN DEVELOPING NERVOUS SYSTEM.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED MATERNALLY AND ZYGOTICALLY.
CC EXPRESSION OF THE MATERNAL TRANSCRIPT DECREASES UNTIL EMBRYONIC
CC STAGE 14, ZYGOTIC TRANSCRIPT IS FIRST DETECTED AT STAGE 11.
CC -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y00274; CAA68377.1; -;
DR EMBL: U17742; AAB60232.1; -;
DR EMBL: AE003680; AAF54270.1; -;
DR PIR: A29253; A29253.
DR TRANSFAC: T00395; -;
DR FLYBASE: FBgn0001180; hb.
DR INTERPRO: IPR000822; -;
DR PFAM: PF00096; zf-C2H2; 6.
DR PRINTS: PR00048; ZINC_FINGER.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 3.
KW Developmental protein; Gap; zinc-finger; Metal-binding; DNA-binding;
KW Repeat; Nuclear protein; Polymorphism.
FT DOMAIN 240 349 ZINC FINGERS I.
FT DOMAIN 705 757 ZINC FINGERS II.
FT DOMAIN 61 69 POLY-GLN.
FT DOMAIN 102 108 POLY-GLN.
FT DOMAIN 117 122 POLY-HIS.
FT ZN_FING 240 262 C2H2-TYPE.
FT ZN_FING 269 291 C2H2-TYPE.
FT ZN_FING 297 319 C2H2-TYPE.
FT ZN_FING 325 349 C2H2-TYPE.
FT DOMAIN 404 409 POLY-GLN.
FT DOMAIN 555 561 POLY-GLN.
FT ZN_FING 705 727 C2H2-TYPE.

FT ZN_FING 733 757 C2H2-TYPE.
FT VARIANT 437 437 A -> P.
FT VARIANT 649 649 V -> M.
SQ SEQUENCE 758 AA; 83113 MW; AD74802EBB56ACD7 CRC64;

Query Match 11.58; Score 94.5; DB 1; Length 758;
Best Local Similarity 24.6%; Pred. No. 1.5;
Matches 43; Conservative 19; Mismatches 56; Indels 57; Gaps 7;
QY 2 CHSRSCH-----PTMTILQAPTAPS---TIPGPRGSGPEIFFDLPPEPAAAPAGR 51
DB 337 CHSFKLHLRKYGHKPGMVLDEGTNPNSLVIDYGTTRG-----PKSKNGG 382
QY 52 PSASRCHRRKRSRRVLYPRVVRQLPVEEPNPAKRLLFLTLTIVFCOILMAEEGVP----- 106
DB 383 PIASGSGSGSRKSNVAAPQO---QOSOPAQ-----PVATLSAALOGFPLVQON 432
QY 107 -----APLPEDAPNAASLAPTPVSPVLEPFPNLTSEPSDYALDLSTFLOQH 152
DB 433 SAPPAASPVLPPLPASPAKSVASVEQTPSLP-----SPANLLPPLASLQON 478

Search completed: January 30, 2001, 23:12:24
Job time: 4005 sec

RT "Molecular cloning of chicken myosin-binding protein (MyBP) H (86-kDa
RT protein) reveals extensive homology with MyBP-C (C-protein) with
RT conserved immunoglobulin C2 and fibronectin type III motifs";
RL J. Biol. Chem. 268:3670-3676(1993).
CC -!- FUNCTION: BINDS TO MYOSIN; PROBABLY INVOLVED IN INTERACTION WITH
CC THICK MYOFILAMENTS IN THE A-BAND.
CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE. SEEMS TO BE ALSO EXPRESSED IN
CC THE SLOW TONIC AID MUSCLE. NOT DETECTED IN GIZZARD OR HEART.
CC -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE MYBP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L05605; AAA21418.1; -
DR INTERPRO: IPR001777; -
DR INTERPRO: IPR002965; -
DR INTERPRO: IPR003006; -
DR PFAM: PF00041; fn3; 2.
DR PFAM: PF00047; ig; 2.
DR PRINTS: PR00014; FNTYPEIII.
DR PRINTS: PR01217; PRICHEXTENS.
KW Immunoglobulin domain; Cell adhesion; Muscle protein; Thick filament;
FT Repeat.
FT DOMAIN 135 221 FIBRONECTIN TYPE-III.
FT DOMAIN 253 312 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 331 416 FIBRONECTIN TYPE-III.
FT DOMAIN 458 518 IG-LIKE C2-TYPE DOMAIN.
FT CONFLICT 2 2 T -> G (IN AA SEQUENCE).
FT CONFLICT 9 9 A -> P (IN AA SEQUENCE).
FT CONFLICT 15 15 A -> K (IN AA SEQUENCE).
FT SEQUENCE 537 AA; 58678 MW; 06C4CF0EF1DD233 CRC64;

Query Match 11.6%; Score 95.5; DB 1; Length 537;
Best Local Similarity 27.5%; Pred. No. 0.89;
Matches 38; Conservative 6; Mismatches 61; Indels 33; Gaps 5;
OY 9 PTMTILQAPTTPAPSTIPGRSGPEITFDPLPEPAAAPAGRPSASRGRKSRRLVLP 68
DB 38 PTFKEGHAPTPEEHAPPKKEHAP-----PKKEHAPAPAAETPPAPCHPDAPQAPAP 92
OY 69 RVVRRQLPVEEPNPAKRLLLTLTVFCQILMAEGVPAPLPDPAPNAASLAPTPVSPV 128
DB 93 AAHAPTPTHEAAPAH-----EEG-PPPAAPAEAP-----APEP----- 125
OY 129 LEFFNLTSFSDYALDLS 146
DB 126 -EPEKPEEPSVPLSLA 142

RESULT 13
XP2_XENLA
ID XP2_XENLA STANDARD: PRT: 439 AA.
AC P17437; Q08944;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEP PROTEIN).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
RN [1]
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A.
RX MEDLINE=92332564; PubMed=1629230;
RA Hauser F., Roeben C., Hoffmann W.;

RT "xp2, a new member of the P-domain peptide family of potential growth
RT factors, is synthesized in Xenopus laevis skin.";
RL J. Biol. Chem. 267:14451-14455(1992).
RN [2]
RP SEQUENCE OF 3-439 FROM N.A.
RC TISSUE=SKIN;
RX MEDLINE=90127399; PubMed=2298293;
RA Gnachl M., Berger H., Thalhammer J., Kreil G.;
RT "dermal glands of Xenopus laevis contain a polypeptide with a highly
RT repetitive amino acid sequence.";
RL FEBS Lett. 260:145-148(1990).
CC -!- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE
CC EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS
CC AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS OF THE PROTEINS ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE APEG FORM INCLUDES A LARGE
CC DOMAIN WITH 33 SEQUENCE REPEATS.
CC -!- TISSUE SPECIFICITY: SKIN.
CC -!- SIMILARITY: CONTAINS 2 P-TYPE (TREFOL) DOMAINS.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392
CC ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M90095; AAA50001.1; -
DR EMBL: X51394; CAA35759.1; ALT_FRAME.
DR PIR: S07498; SKXLG.
DR PIR: A37331; A37331.
DR HSSP: P04155; 1PS2.
DR INTERPRO: IPR000519; -
DR PFAM: PF00088; trefol; 2.
DR PRINTS: PR00680; PTREFOL.
DR PROSITE: PS00025; P_TREFOL; 2.
KW Signal; Growth factor; Alternative splicing; Amphibian skin; Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.
FT MOD_RES 23 23 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT DOMAIN 26 343 33 X REPEATS OF G-[GE]-[AP](2,4)-A-E.
FT DOMAIN 350 391 P-TYPE 1.
FT DOMAIN 397 438 P-TYPE 2.
FT DISULFID 351 377 BY SIMILARITY.
FT DISULFID 361 376 BY SIMILARITY.
FT DISULFID 371 388 BY SIMILARITY.
FT DISULFID 398 424 BY SIMILARITY.
FT DISULFID 408 423 BY SIMILARITY.
FT DISULFID 418 435 BY SIMILARITY.
FT VARSPIC 26 343 MISSING (IN ISOFORM XP2).
FT CONFLICT 3 3 H -> S (IN REF. 2).
FT CONFLICT 18 18 C -> W (IN REF. 2).
FT SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;

Query Match 11.5%; Score 95; DB 1; Length 439;
Best Local Similarity 25.7%; Pred. No. 0.8;
Matches 35; Conservative 6; Mismatches 57; Indels 38; Gaps 5;
OY 15 QAPTAP-----STIFGRKSGPEITFDPLPEPAAAPAGRPSASRGRKSRRLVLP 70
DB 247 EAPAPAPAEAPAPAPAEAPAEAPAEAPAEAPAEAPAEAPAEAPAEAPAEAPAEAP 301
OY 71 VRRQLPVE-----EPNPAKRLLLTLTVFCQILMAEGVPAPLPDPAPNAASLAPTP 124
DB 302 AEGEAPAPAPAEAGGAPSP-----AEGGAPAAAEAGGAPAPAPAE 343
OY 125 VSPVLE-----PFNLT 135

```
FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE 620 AA: 65406 MW: 641DD2278AB28524 CRC64;

Query Match 12.2%; Score 100; DB 1; Length 620;
Best Local Similarity 24.3%; Pred. No. 0.46; 75; Indels 44; Gaps 6;
Matches 42; Conservative 12; Mismatches 12;

QY 4 SRSCHTMTILQAPTAPSTIPGPRGSGEIFTFDPLPEPAAAGAPRASGRHRSR 63
DB 173 SRHGHPPPTTAQPTPIYSPQVQPP---TYSPPPTHVQTPSP-PSRGHQPPQ 228
QY 64 RVLY-PRVVRRLPVEENPAKRLLELLTIVFCQILMAEGVPAP-----LPED 113
DB 229 TRHAPPTHRRAPTHQPSPLHL-----PSPRRQPPPTYSPPPP 270
QY 114 A-----PNAASLAPTPVSPVLEPFLNLTSEPSDYALDLSTELQHPA 154
DB 271 AYAQSPQSPSTYSPPPTYSPPPPSPPIYSPPPPAYSPGPTPTTFSPPPPA 323

RESULT 10
Y091.NPVO
ID Y091.NPVO STANDARD; PRT; 279 AA.
AC O10341;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOTHETICAL 29.3 KDA PROTEIN (ORF92).
OS Orygia pseudotsugata multicapsid polyhedrosis virus (OpMPNV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
NC Nucleopolyhedrovirus.
CC [1]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC
DR EMBL; X14112; CAA32311.1; -
KW Repeat.
FT DOMAIN 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q.
SQ SEQUENCE 3164 AA: 335857 MW: CC5D31FF4F9FE3F4 CRC64;

Query Match 12.0%; Score 98.5; DB 1; Length 3164;
Best Local Similarity 27.3%; Pred. No. 3.1;
Matches 47; Conservative 7; Mismatches 57; Indels 61; Gaps 8;

QY 2 CHRSCHTMTILQ--PTPAPSTIPG-----RRSGPEIFTFDPLPEPAAAP----- 48
DB 2612 CAPRPPGPAVTAAREARPGVPAESTRAPVGRDPRRLPSQ---SSPAPDPATAPRPPA 2668
QY 49 -----AGRPSASRGHRKRSRVL-----YPRVV-----RRQLPVEE 79
DB 2669 SSRASASSSSGRARRRRARRSLARATOAATTOGWRPPALPDTPVPTDFARPAPPKP 2728
QY 80 PNPARKRLLELLTIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLEP 131
DB 2729 PEPAPHALV-----SGVPLPLGPPQAAGQAS--PALPIDPVPPP 2764

RESULT 12
MYPH.CHICK
ID MYPH.CHICK STANDARD; PRT; 537 AA.
AC Q05623;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MYOSIN-BINDING PROTEIN H (MYBP-H) (86 KDA PROTEIN).
GN MYBP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CC [1]
CC SEQUENCE FROM N.A., AND SEQUENCE OF 2-37.
RN RP
PC TISSUE-PECTORALIS MUSCLE;
RX MEDLINE-93155224; PubMed-7679114;
RA Vaughan K.T., Weber F.E., Einheber S., Fischman D.A.;
```

```
TEGU_HSV11
ID TEGU_HSV11 STANDARD; PRT; 3164 AA.
AC P10220;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE LARGE TEGUMENT PROTEIN (VIRION PROTEIN UL36).
GN UL36.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
CC [1]
CC SEQUENCE FROM N.A.
RX MEDLINE-88274327; PubMed-2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1988).
CC -!- SIMILARITY: TEGUMENT PROTEIN.
CC -!- FUNCTION: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC
DR EMBL; X14112; CAA32311.1; -
KW Repeat.
FT DOMAIN 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q.
SQ SEQUENCE 3164 AA: 335857 MW: CC5D31FF4F9FE3F4 CRC64;

Query Match 12.0%; Score 98.5; DB 1; Length 3164;
Best Local Similarity 27.3%; Pred. No. 3.1;
Matches 47; Conservative 7; Mismatches 57; Indels 61; Gaps 8;

QY 2 CHRSCHTMTILQ--PTPAPSTIPG-----RRSGPEIFTFDPLPEPAAAP----- 48
DB 2612 CAPRPPGPAVTAAREARPGVPAESTRAPVGRDPRRLPSQ---SSPAPDPATAPRPPA 2668
QY 49 -----AGRPSASRGHRKRSRVL-----YPRVV-----RRQLPVEE 79
DB 2669 SSRASASSSSGRARRRRARRSLARATOAATTOGWRPPALPDTPVPTDFARPAPPKP 2728
QY 80 PNPARKRLLELLTIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLEP 131
DB 2729 PEPAPHALV-----SGVPLPLGPPQAAGQAS--PALPIDPVPPP 2764

RESULT 12
MYPH.CHICK
ID MYPH.CHICK STANDARD; PRT; 537 AA.
AC Q05623;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MYOSIN-BINDING PROTEIN H (MYBP-H) (86 KDA PROTEIN).
GN MYBP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CC [1]
CC SEQUENCE FROM N.A., AND SEQUENCE OF 2-37.
RN RP
PC TISSUE-PECTORALIS MUSCLE;
RX MEDLINE-93155224; PubMed-7679114;
RA Vaughan K.T., Weber F.E., Einheber S., Fischman D.A.;
```

DR PROSITE; PS50014; BROMODOMAIN_2: 1.
KW Proto-oncogene; Chromosomal translocation; DNA-binding;
KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
KW Alternative splicing; Polymorphism.
FT NON_TER 1 1
FT DNA_BIND 67 78 A..T HOOK (BY SIMILARITY).
FT DNA_BIND 115 125 A..T HOOK (BY SIMILARITY).
FT DNA_BIND 199 207 A..T HOOK (BY SIMILARITY).
FT DOMAIN 1330 1381 PHD 1.
FT DOMAIN 1383 1432 PHD 2.
FT DOMAIN 1465 1529 PHD 3.
FT DOMAIN 1605 1650 BROMODOMAIN (DIVERGENT).
FT DOMAIN 3737 3846 SET DOMAIN.
FT DOMAIN 35 41 POLY-GLY.
FT DOMAIN 459 469 PRO-RICH.
FT DOMAIN 1231 1238 POLY-PRO.
FT DOMAIN 3533 3536 POLY-GLU.
FT DOMAIN 3693 3697 POLY-GLU.
FT VARSPLIC 1503 1505 MISSING (IN ISOFORM 2).
FT VARIANT 1497 1497 K -> T.
SQ SEQUENCE 3866 AA; 420976 MW; ADFC55E14E806F1D CRC64;

Query Match 12.5%; Score 102.5; DB 1; Length 3866;
Best Local Similarity 26.8%; Pred. No. 1.8;
Matches 41; Conservative 19; Mismatches 48; Indels 45; Gaps 8;

OY 12 TILQAP-TPAPSTIPGRRGSGPEIFTFDLPLE-----PAAAPAGRPES--ASR 56
| : : : | | | | | : : : | | | | : :
DB 1129 TSVKSPLEPAQKAAPPPREAPKSSSEPPRPKPVKESEEGAPAPAPAPKQVSAP 1188
| : : : | | | | | : : : | | | | : :
OY 57 GHRKRRRLVPRVRRQLPVEENPAKRLFLLLTVFCQILMAEGVPA-----PLP 110
| : : : | | | | | : : : | | | | : :
DB 1189 ASRKSQKQVSPAAV--VPPQPPSTAP-----QKKEAPKAVPSEPKKKQPPP 1233
| : : : | | | | | : : : | | | | : :
OY 111 PEDAPNAA---SLAPTPVSPVLEPFLNTSEPSD 140
| : : : | | | | | : : : | | | | : :
DB 1234 PEPGEQSKQKQKVAPLSPVQK-----KPKD 1260
| : : : | | | | | : : : | | | | : :

RESULT 8
NO20_MEDTR STANDARD; PRT; 268 AA.
AC P93329;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE EARLY NODULIN 20 PRECURSOR (N-20).
GN ENOD20.
OS Medicago truncatula (Barrel medic).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids I;
OC Fabales; Fabaceae; Papilionoideae; Medicago.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. JEMALONG;
RX MEDLINE=98187258; PubMed=9526510;
RA Greene E.A., Erard M., Dedieu A., Barker D.G.;
RT "MENOD16 and 20 are members of a family of phytoecyanin-related early nodulins".
RL Plant Mol. Biol. 36:775-783(1998).
CC -!- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X99467; CAA67830.1; -
KW Nodulation; Signal.

FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 268 EARLY NODULIN 20.
FT DOMAIN 23 7 PLASTOCYANIN-LIKE.
FT DOMAIN 136 145 POLY-PRO.
SQ SEQUENCE 268 AA; 28668 MW; 6FA29C5798C75F91 CRC64;

Query Match 12.3%; Score 101.5; DB 1; Length 268;
Best Local Similarity 28.3%; Pred. No. 0.15;
Matches 41; Conservative 12; Mismatches 31; Indels 61; Gaps 9;

OY 13 ILQAPTAPPS-----TIPGRRGSGPEIFTFDLPPEAAAPAGRPESASGRHKRRSR 64
| : : : | | | | | : : : | | | | : :
DB 132 VLSSPPPPSPPTPRSTPIPHPRSLPS-----PSPSPSPSPSPSPSPSPSPSP 178
| : : : | | | | | : : : | | | | : :
OY 65 VLYPRVRRQLPVEENPAKRLFLLLTVFCQILMAEGVPAPLP-----PEDAPNA 117
| : : : | | | | | : : : | | | | : :
DB 179 -----PRSTPI--PHPRKR-----SPASPSPSLSKSPSPSPSPSPSPSP 211
| : : : | | | | | : : : | | | | : :
OY 118 ASLAPTPVSPV--LEPFLNTSEPSD 140
| : : : | | | | | : : : | | | | : :
DB 212 -SLAPSPSDSVASLAP---SSSPSD 232
| : : : | | | | | : : : | | | | : :

RESULT 9
EXTN_TOBAC STANDARD; PRT; 620 AA.
AC P13983;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
GN HRGPNT3.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. XANTHI; TISSUE=LEAF;
RX MEDLINE=90128263; PubMed=2612909;
RA Keller B., Lamb C.J.;
RT "Specific expression of a novel cell wall hydroxyproline-rich
RT glycoprotein gene in lateral root initiation".
RL Genes Dev. 3:1639-1646(1989).
CC -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC MAIN ROOT.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X13885; CAA32090.1; -
DR PIR; S06733; S06733.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
CC Hydroxylation.
FT CHAIN 1 2 ? EXTENSIN.
FT CHAIN 70 73 H-A-P-P.
FT REPEAT 148 151 H-A-P-P.
FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT REPEAT 229 235 1.
FT REPEAT 236 242 2.
FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.

FT ACT_SITE 181 181 BY SIMILARITY.
FT MOD_RES 218 218 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT 220 220 (BY SIMILARITY).
FT MOD_RES 220 220 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT 337 340 (BY SIMILARITY).
FT DOMAIN 433 464 POLY-ALA.
FT DOMAIN 520 523 PRO-RICH 1.
FT DOMAIN 577 699 POLY-ARG.
FT CONFLICT 19 32 PRO-RICH 2.
FT AREGTRPHRLCS -> GPKVPEPAHTAASVA (IN
REF. 2).
FT CONFLICT 609 609 L -> V (IN REF. 2).
FT SEQUENCE 815 AA; 88636 MW; 379AD69803207CCF CRC64;
Query Match 12.6%; Score 104; DB 1; Length 815;
Best Local Similarity 24.2%; Pred. No. 0.29;
Matches 48; Conservative 23; Mismatches 57; Indels 70; Gaps 10;
QY 4 SRSCPTMTILQAPTPAPSTIPGP-----RRSGPEI-----FTFDPLPEPAAP 48
Db 575 ARPAAPALTSVPAPAPAPTPTPTVQPTSPPGPLAQTPGQPSAGSTSGVPQPACPP 634
QY 49 AG---RPSASRGHRKRSRVLYPRVVRQLPVEEPNPAKRLLLLTIVFCOILMAEGV 105
Db 635 PGAPHTPGPG-----PI--PVPAPPQIATSTSLAAGSLVPPPG 674
QY 106 PAP-----LPPEDAPNA--ASLAPTP-----VSPVLEP-FNLTSE 137
Db 675 PGSSPTGVLPPFPGLPPDPDAGCAPQSSSESDVNLVTOOLSKSQVEDPLPPVSGTPK 734
QY 138 PS-----DYALDSTFLQ 151
Db 735 GSGAGYGVGFDEELFNQ 752
RESULT 6
MODD_MYCAV STANDARD; PRT; 381 AA.
AC Q48919;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MOLYBDATE UPTAKE SECRETED PROTEIN PRECURSOR (FIBRONECTIN ATTACHMENT
DE PROTEIN) (FAP-A).
GN MODD.
OS Mycobacterium avium.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=101;
RX MEDLINE=97011577; PubMed=8858587;
RA Schorey J.S., Holsti M.A., Ratliff T.L., Allen P.M., Brown E.J.;
RT "Characterization of the fibronectin-attachment protein of
RT Mycobacterium avium reveals a fibronectin-binding motif conserved
RT among mycobacteria";
RL Mol. Microbiol. 21:321-329(1996).
CC MODABC FOR MOLYBDENUM (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO THE MODD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U53585; AAB50543.1;
DR Molybdenum; Transport; Signal.
KW

FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 381 MOLYBDATE UPTAKE SECRETED PROTEIN.
SQ SEQUENCE 381 AA; 38210 MW; 0DACB9A22AA11D2B CRC64;
Query Match 12.5%; Score 102.5; DB 1; Length 381;
Best Local Similarity 30.2%; Pred. No. 0.18;
Matches 42; Conservative 8; Mismatches 42; Indels 47; Gaps 8;
QY 15 QAPTAP---STIPGRGSGPEIFTFDPLPEPA--AAPAGRPS--ASRGHRKRSRVLYPR 69
Db 42 EYPTVPSTATTTPRRRRIP-----GQADNAAGAPAPAPNCGQORPRR----- 88
QY 70 VVRQLPVEEPNPAKRLLLLTIVFCOILMAEGVPAPLPEDPAPNAASLAPTPVSPVL 129
Db 89 --RRMIPTRAPPPA-----GAPPNGAPPAAPNGA--PPPPVDPNA 124
QY 130 EP-----FNLTSEPSDYAL 143
Db 125 PPPPADPNAGRIPNSYVL 143
RESULT 7
HRX_MOUSE STANDARD; PRT; 3866 AA.
ID HRX_MOUSE
AC P55200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ZINC FINGER PROTEIN HRX (ALL-1) (FRAGMENT).
GN MLL OR HRX OR ALL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX STRAIN=C57BL/6J, AND B6/CBA; TISSUE=SPLEEN, AND LUNG;
RX MEDLINE=93317679; PubMed=8327517;
RA Ma Q., Alder H., Nelson K.K., Chatterjee D., Gu Y., Nakamura T.,
RA Canaan E., Croce C.M., Siracusa L.D., Buchberg A.M.;
RT "Analysis of the murine All-1 gene reveals conserved domains with
RT human ALL-1 and identifies a motif shared with DNA
RT methyltransferases";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).
CC -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC -!- MAY REGULATE GENES INVOLVED IN SKELETAL FORMATION DURING
CC EMBRYONIC DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 PHD-FINGER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L17069; AAA62593.1;
DR MGD; MGI:96995; MLL.
DR INTERPRO: IPR001214;
DR INTERPRO: IPR001487;
DR INTERPRO: IPR001965;
DR INTERPRO: IPR002857;
DR PFAM: PF00628; PHD; 3.
DR PFAM: PF00856; SET; 1.
DR PFAM: PF02008; ZF-CXXC; 1.

9	PTWILQAPTPAS	-----TIGPRRG-SGPEIFTFDPLPEPAAAPAGRSASGRHKRS	62
484	PVSTIAPSTPSRLQIPILPQAAASNPKILPTTSPSPTAAA	-----PTTT-----	534
63	RRVLYPRVVRQ	-----LPVEEPNPKRLLELLITVFCOILMAEGVPAP	108
535	-TLSPSPHIVQQPQPSAAPAPSPLLPQQQPTPS	-----AAPAPSP	572
109	LPPEADAPNAASLAPTPVSPVLEPNLTYS	-----EPSDYALDLSTFLQ-----QHPAA	155

FT	BINDING	83
FT	BINDING	83

[illegible]

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: S81914; AAB36278.1; -;
EMBL: X96438; CAA65304.1; -;
EMBL: Y14551; CAA74886.1; -;
EMBL: T49531; -; NOT_ANNOTATED_CDS.
MIM: 602996; -;
KW Glycoprotein; Transmembrane; Signal-anchor.
FT DOMAIN 1 82 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 83 99 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT DOMAIN 100 156 * EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 54 54 A -> G (IN REF. 1).
FT CONFLICT 106 106 P -> R (IN REF. 1).
SQ SEQUENCE 156 AA; 16928 MW; 83C06116C80B9240 CRC64;

Query Match 100.0%; Score 823; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.6e-58;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASL 120
QY 121 APTVPVLEPNTSEPSDYALDLSLTFLOQHPAAF 156
DB 121 APTVPVLEPNTSEPSDYALDLSLTFLOQHPAAF 156

RESULT 2
ID IEX1_MOUSE STANDARD; PRT; 153 AA.
AC P46694;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN).
GN IER3 OR IEX1 OR GLY96.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=93173526; PubMed=8437864;
RA Charles C.H., Yoon J.K., Simske J.S., Lau L.F.;
RT "Genomic structure, cDNA sequence, and expression of gly96, a growth factor-inducible immediate-early gene encoding a short-lived glycosylated protein.";
RL Oncogene 8:797-801(1993).
CC -!- FUNCTION: NOT KNOWN; EXPRESSED DURING THE G0-G1 TRANSITION OF THE CELL CYCLE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE LUNG, TESTES AND THE UTERUS.
CC -!- INDUCTION: BY SERUM GROWTH FACTORS AND TPA.
CC -!- PTM: GLYCOSYLATED.
CC -!- SIMILARITY: STRONG, TO HUMAN ORTHOLOG.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X67644; -; NOT_ANNOTATED_CDS.
PIR: S33363; S33363.
MGD: MGI:104814; IER3.
KW Glycoprotein; Transmembrane; Signal-anchor.
FT DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 86 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT DOMAIN 103 153 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 153 AA; 16875 MW; F4429E3B0120DE11 CRC64;

Query Match 63.8%; Score 525; DB 1; Length 153;
Best Local Similarity 70.3%; Pred. No. 4.2e-35;
Matches 109; Conservative 9; Mismatches 29; Indels 8; Gaps 2;

QY 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASL 120
QY 121 APTVPVLEPNTSEPSDYALDLSLTFLOQHPAAF 153
DB 121 APTVPVLEPNTSEPSDYALDLSLTFLOQHPAAF 153

RESULT 3
ID IEX1_MOUSE STANDARD; PRT; 283 AA.
AC P24152;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE EXTENSIN PRECURSOR (PROLINE-RICH GLYCOPROTEIN).
GN HRGP.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RX MEDLINE=91370882; PubMed=1893107;
RA Raz R., Cretin C., Puigdomenech P., Martinez-Izquierdo J.A.;
RT "The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum vulgare.";
RL Plant Mol. Biol. 16:365-367(1991).
CC -!- FUNCTION: STRUCTURAL COMPONENT IN PRIMARY CELL WALL.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN GLYCOSYLATED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X56010; CAA39485.1; -;
PIR: S14449; S14449.
INTERPRO: IPR002965;
DR PRINTS: PR01217; PRICHEXTENSIN
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein; Hydroxylation.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:05:39 ; Search time 43.61 Seconds
(without alignments)
115.521 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILQAPTPA.....EPSDYALDLSTFLQHPAAAF 156

Scoring table: BLOSUM62 *
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	823	100.0	156	1	P46695 homo sapien
2	525	63.8	153	1	IEXL_MOUSE
3	112.5	13.7	283	1	EXTN_SORBI
4	105	12.8	3149	1	TEGU_EBV
5	104	12.6	815	1	Q13164 homo sapien
6	102.5	12.5	381	1	MODD_MYCAV
7	102.5	12.5	3866	1	HRX_MOUSE
8	101.5	12.3	268	1	NO20_MEDTR
9	100	12.2	620	1	EXTN_TOBAC
10	99.5	12.1	279	1	Y091_NPVOP
11	98.5	12.0	3164	1	TEGU_HSV11
12	95.5	11.6	537	1	MYPH_CHICK
13	95	11.5	439	1	XP2_XENLA
14	94.5	11.5	510	1	ERK_HUMAN
15	94.5	11.5	758	1	HUNB_DROME
16	94.5	11.5	1595	1	SOS_DROME
17	94	11.4	1290	1	PER1_HUMAN
18	93.5	11.4	1402	1	IF4G_RABIT
19	93	11.3	817	1	VRPL_YEAST
20	92.5	11.2	605	1	BRL1_EBV
21	92.5	11.2	625	1	NIFA_AZOBR
22	92	11.2	534	1	APG_ARATH
23	92	11.2	3969	1	HRX_HUMAN
24	91.5	11.1	267	1	EXTN_MAIZE
25	91.5	11.1	615	1	MUTL_ECOLI
26	91.5	11.1	1133	1	SREL_CRIGR
27	90	10.9	316	1	CDCN_HUMAN
28	90	10.9	449	1	APG_BRANA
29	89.5	10.9	1487	1	ICP4_HSVK
30	89.5	10.9	2715	1	TRX2_HUMAN
31	89	10.8	1291	1	PER1_MOUSE
32	88.5	10.8	865	1	CPN_DROME
33	88	10.7	474	1	VTP3_TTV1V

34	88	10.7	477	1	BLAR_HUMAN	P08588 homo sapien
35	87.5	10.6	543	1	P461_NPVAC	Q03209 autographa
36	87.5	10.6	591	1	MNT_MOUSE	Q08789 mus musculus
37	87	10.6	1575	1	SYJ1_HUMAN	Q43426 homo sapien
38	87	10.6	2414	1	P300_HUMAN	Q09472 homo sapien
39	86.5	10.5	1023	1	SREL_RAT	P56720 rattus norv
40	86.5	10.5	1487	1	ICP4_HSVB	P28925 equine herp
41	86	10.4	390	1	VGJ1_HSV11	Q9Y4h2 herpes simp
42	86	10.4	1324	1	IRS2_HUMAN	P06487 herpes simp
43	85.5	10.4	224	1	Y091_NPVAC	Q9Y4h2 homo sapien
44	85.5	10.4	626	1	GPBA_HUMAN	P41479 autographa
45	85.5	10.4	1233	1	NME3_HUMAN	P07359 homo sapien
						Q14957 homo sapien

ALIGNMENTS

RESULT 1	
IEX1_HUMAN	
ID	IEX1_HUMAN STANDARD; PRT; 156 AA.
AC	P46695; Q93044; Q92691;
DT	01-NOV-1995 (Rel. 32, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY
DE	PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN) (PACAP-RESPONSIVE
DE	GENE 1 PROTEIN) (PRGI PROTEIN) (DIFFERENTIATION-DEPENDENT GENE 2
DE	PROTEIN) (DIF-2 PROTEIN)
GN	IER3 OR IEX1 OR PRG1 OR DIF2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PLACENTA;
RX	MEDLINE=96181295; PubMed=8603392;
RA	Kondratyev A.D., Chung K.-N., Jung M.O.;
RT	"Identification and characterization of a radiation-inducible
RT	glycosylated human early-response gene.";
RL	Cancer Res. 56:1498-1502(1996).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=96221139; PubMed=8653710;
RA	Schaefer H., Trauzold A., Siegel E.G., Folsch U.R., Schmidt W.E.;
RT	"PRG1: a novel early-response gene transcriptionally induced by
RT	pituitary adenylate cyclase activating polypeptide in a pancreatic
RT	carcinoma cell line.";
RL	Cancer Res. 56:2641-2641(1996).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=97339426; PubMed=9196025;
RA	Pietzsch A., Buechler C., Aslanidis C., Schmitz G.;
RT	"Identification and characterization of a novel monocyte/macrophage
RT	differentiation-dependent gene that is responsive to
RT	lipopolysaccharide, ceramide, and lysophosphatidylcholine.";
RL	Biochem. Biophys. Res. Commun. 235:4-9(1997).
RN	[4]
RP	PRELIMINARY SEQUENCE OF 1-106 FROM N.A.
RC	TISSUE=PLACENTA;
RA	Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M.,
RA	Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,
RA	Rifkin L., Rohling T., Tan F., Trevaskis E., Waterston R.,
RA	Williamson A., Wohldmann P., Wilson R.;
RT	Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC	-1- INDUCTION: BY RADIATION, TPA, OKADAIC ACID AND TNF-ALPHA.
CC	-1- PTM: GLYCOSYLATED.
CC	-1- SIMILARITY: STRONG, TO MOUSE ORTHOLOG.
CC	-1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC	FRAMESHIFTS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-496 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96773.1
A:Experimental source: specific host *Chlorella* strain NC64A
C:Genetics:
A:Note: A405R

Query Match 13.0%; Score 107; DB 2; Length 496;
Best Local Similarity 22.8%; Pred. No. 0.19;
Matches 37; Conservative 20; Mismatches 57; Indels 48; Gaps 6;
QY 16 APTAPSTIPGPRGSGPEIFTDPLPEAAAPAGRSASRGHKKRRVLYPRVVRQL 75
DB 128 APKAPKAPKAPKSPK*----PAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 180
QY 76 PVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLP-----PEDAPNAASLAP----- 122
DB 181 PKPAKPAK*-----PAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 226
QY 123 -----TPVSPVLE-----PFLTSEPSDYALDLSTFLQ 150
DB 227 DINDKAPMLPIEDIPSSFOOTFTVTSITTTATNIDLPGYVK 268

RESULT 7
A49227
sialidase - Actinomyces viscosus
C:Species: Actinomyces viscosus
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 22-Oct-1999
C:Accession: A49227
R:Yeung, M.K.
Infect. Immun. 61, 109-116, 1993
A:Title: Complete nucleotide sequence of the Actinomyces viscosus T14V sialidase gene: P
A:Reference number: A49227; MUID:93114861
A:Contents: T14V
A:Accession: A49227
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-901 <YEU>
A:Cross-references: GB:L06898; NID:g289087; PIDN:AAA21932.1; PID:g141852
A:Note: sequence extracted from NCBI backbone (NCBI:121598, NCBI:P:121599)

Query Match 12.9%; Score 106.5; DB 2; Length 901;
Best Local Similarity 29.0%; Pred. No. 0.37;
Matches 45; Conservative 19; Mismatches 57; Indels 27; Gaps 8;
QY 9 PTWTILOAPYPAVSTIPGPRGSGPEIFTDPLPEAAAPAGRSASRGHKKRRVLYP 68
DB 722 PTTAPSTERTPAAPSAPEQTDGP---TAAPAPETSSAPAAEPTQA-----PTVAP 770
QY 69 RVVRRQIPVEEPNPAKR--LFLLLTIVFCQILMA--EGVP-----APLPEDAPNAASL 120
DB 771 SVPTQAGQAPGAPKPGATGRAPSVNPKATGAATEPTGPTSSASAPSRNA----- 824
QY 121 APTPVSPVLEPFLNLTSEPSDYALDLSTFLQHPAA 155
DB 825 APTP-KPGMEPDEI-DRPSDGTMAOPTGGASAPSA 857

RESULT 8
T00037
hypothetical protein KIAA0284 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C:Accession: T00037
R:Ohara, O.; Nagase, T.; Ishikawa, K.; Nakajima, D.; Ohira, M.; Seki, N.; Nomura, N.
submitted to the EMBL Data Library, August 1997
A:Description: Prediction of the coding sequences of unidentified human genes.
A:Reference number: 214073
A:Accession: T00037

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1323 <OHA>
A:Cross-references: EMBL:AB006622; NID:d1170675; PIDN:BAA22953.1; PID:d1C23829
A:Experimental source: brain
C:Genetics:
A:Note: KIAA0284

Query Match 12.9%; Score 106.5; DB 2; Length 1323;
Best Local Similarity 26.8%; Pred. No. 0.54;
Matches 45; Conservative 13; Mismatches 67; Indels 43; Gaps 7;
QY 3 HSRSCHPTMTILOAPTTPA-----PSTIPGPRGSGPEIF-----TFDPL--PEPAA 46
DB 368 HSPSGDPKADKRRGPTPADRDSPVPAPVAGSGSPQAGSLKREKTEERLGSPSPAS 427
QY 47 APAGRSASRGHKKRRVLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVP 106
DB 428 RTPARFGSVGRKSLAQDFMAOCLRESSPAARPSPEK-----VP 467
QY 107 APLPPEDAPNAASLA--PTPVSPVLEPFLNLT-----SEPSDYALDLSTF 148
DB 468 PVLPAPLTHGTSVPGPTPPAPTDP-QLTKARKQEBDDSLSDAGTY 514

RESULT 9
T30351
mucin-like protein - Lymantria dispar nuclear polyhedrosis virus
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T30351
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.;
Virology 253, 17-34, 1999
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lym
A:Reference number: 220836; MUID:99124785
A:Accession: T30351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1029 <KUZ>
A:Cross-references: EMBL:AF081810; PIDN:AAC70189.1

Query Match 12.9%; Score 106; DB 2; Length 1029;
Best Local Similarity 29.5%; Pred. No. 0.46;
Matches 51; Conservative 6; Mismatches 52; Indels 64; Gaps 8;
QY 16 APTAPSTIPGPRGSGP-EIETFF-----DPLPEAAAPAGRSASRCHKRSRRLV 67
DB 637 APEAPEPVPEP---SAPVDYFTLSAEPAPEPAPEPAPEPAPEPAPEPMTSRTSEPALG 693
QY 68 PRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSP 127
DB 694 P-VEPALEPVE-----PALEPVEPALGPGPEPGE 725
QY 128 VLEPNLNSE-----PSDYALDL-----STF-----LQHPAA 155
DB 726 ALEPGESTAAASELAVERPAQAPDLTSDSAVESTFGHKSTVASELAARPA 778

RESULT 10
S50754
hypothetical protein WP6 - Chlamydomonas eugametos
C:Species: Chlamydomonas eugametos
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C:Accession: S50754
R:Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.;
Plant Mol. Biol. 26, 947-960, 1994
A:Title: Domain conservation in several volvocalean cell wall proteins.
A:Reference number: S50754; MUID:95093034
A:Accession: S50754
A:Status: preliminary
A:Molecule type: mRNA

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: January 30, 2001, 21:28:35 ; Search time 21.17 Seconds
(without alignments)
500.354 Million cell updates/sec

Title: US-08-799-910-10
Perfect score: 823
Sequence: 1 MCHSRSCHPTMTILOAPTPA.....EPSDYALDLSTFLQOHPAAF 156
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	819	99.5	156	2 JC5537	differentiation-de
2	525	63.8	153	2 S33363	gly96 protein - mo
3	112.5	13.7	283	2 S13383	hydroxyproline-ric
4	109.5	13.3	552	2 T08148	proline-rich myros
5	107.5	13.1	581	2 T22341	hypothetical prote
6	107	13.0	496	2 T17908	proline/lysine-ric
7	106.5	12.9	901	2 A49227	sialidase - Actino
8	106.5	12.9	1323	2 T00037	hypothetical prote
9	106	12.9	1029	2 T30351	mucin-like protein
10	105	12.8	351	2 S50754	hypothetical prote
11	105	12.8	549	2 T17525	proline-rich prote
12	105	12.8	3149	1 Q0BE8	BPLFI protein - hu
13	104.5	12.7	535	1 S76953	protein kinase (EC
14	104	12.6	815	2 B56708	extracellular sign
15	103	12.5	241	2 T22216	hypothetical prote
16	102.5	12.5	288	2 T17737	proline-rich prote
17	102.5	12.5	3866	2 B48205	All-1 protein -GTE
18	102.5	12.5	3869	2 A48205	All-1 protein +GTE
19	101	12.3	339	2 T17636	proline-rich prote
20	100	12.2	620	2 S06733	hydroxyproline-ric
21	99.5	12.1	279	2 T10361	hypothetical prote
22	98.5	12.0	3164	1 WMBEH6	UL36 protein - hum
23	98	11.9	473	2 S50755	hypothetical prote
24	97.5	11.8	225	2 T17815	proline-rich prote
25	97.5	11.8	2157	2 S71461	proline-rich prote
26	97.5	11.8	3938	2 T42761	Bassoon protein -
27	97	11.8	461	2 T10741	extensin-like prot
28	96.5	11.7	168	2 S52994	arabinogalactan-li
29	96	11.7	185	2 T00519	proline-rich prote

30	95.5	11.6	524	2 A75588	probable protein k
31	95.5	11.6	537	2 A46611	myosin-binding pro
32	95.5	11.6	753	2 JQ0532	OP protein - Ksne
33	95.5	11.6	1596	2 A41216	guanine nucleotide
34	95	11.5	202	2 JQ0964	hydroxyproline-ric
35	95	11.5	377	2 A48018	mucin 7 precursor,
36	95	11.5	416	1 SKXLAG	dermal gland prote
37	95	11.5	801	2 T29018	hypothetical prote
38	95	11.5	3511	2 A59295	unconventional myo
39	94.5	11.5	510	1 S43692	transcription fact
40	94.5	11.5	758	2 A29253	finger protein hun
41	94.5	11.5	2342	2 T13412	hypothetical prote
42	94	11.4	328	2 JQ0985	hydroxyproline-ric
43	94	11.4	925	2 S27920	nuclear antigen EB
44	94	11.4	1290	2 T00018	period protein hom
45	93.5	11.4	316	2 T31880	hypothetical prote

ALIGNMENTS

RESULT 1
JC5537
differentiation-dependent protein DIF-2 - human
C:Species: Homo sapiens (man)
C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 07-May-1999
C:Accession: JC5537
R:Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.
Biochem. Biophys. Res. Commun. 235, 4-9, 1997
A:Title: Identification and characterization of a novel monocyte/macrophage diff
A:Reference number: JC5537; MUID:97339426
A:Accession: JC5537
A:Molecule type: mRNA
A:Residues: 1-156 <PIF>
A:Experimental source: monocyte
A>Note: the authors translated the codon CCG for residue 106 as Arg

Query Match 99.5%; Score 819; DB 2; Length 156;
Best Local Similarity 99.4%; Pred. No. 1.2e-58;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MCHSRCHPTMTILOAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
|||||
DB 1 MCHSRCHPTMTILOAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEECVAPLPPEPAPNAASL 120
|||||
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEPAPNAASL 120
QY 121 APTPVSPVLPEPFLNLTSEPSDYALDLSTFLQOHPAAF 156
|||||
DB 121 APTPVSPVLPEPFLNLTSEPSDYALDLSTFLQOHPAAF 156

RESULT 2
S33363
gly96 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C:Accession: S33363
R:Charles, C.H.; Yoon, J.K.; Simske, J.S.; Lau, L.F.
Oncogene 8, 797-801, 1993
A:Title: Genomic structure, cDNA sequence, and expression of gly96, a growth fac
A:Reference number: S33363; MUID:93173526
A:Accession: S33363
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-153 <CHA>
A:Cross-references: EMBL:X67644
C:Genetics:
A:Introns: 70/3
C:Keywords: transmembrane protein


```

Query Match      13.1%; Score 107.5; DB 5; Length 581;
Best Local Similarity 25.3%; Pred. No. 0.033;
Matches 42; Conservative 9; Mismatches 60; Indels 55; Gaps 6;

QY 9 PTWTILQAPTPAPSTIPGPRRSGSPEIFTFDPLPEPAAAP-----AGRPSASR 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 PPADAAPAPPPAPEPVA-----PAPAPEAAPVAPSADAGGYAAAAAPAGG 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 57 GHRKRRRV-----LYPRVVRRLQVPVEPNPAKRLLFLLLTIVFCQILMAEEGVPA 108
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 SYPAKKRRVARDYAEGEAPAAAPAEAPAPAPAPAPA-----PEAAPAPEPAPAP 509
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 109 LP-----PEDAPNAASLAPTIV-----SPVLEPNLTSEPSDYA 142
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 510 APAPAEAPAAAPADAAAPAEAPVPEVAPAPAPAPAPAPAPAPESDASGYS 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: January 30, 2001, 23:11:29
Job time: 3994 sec

```

```
RX MEDLINE=97127825; PubMed=8972576;
RA Ramon A., Gil R., Bursal M., Sentandreu R., Valentin E.;
RT "A novel cell wall protein specific to the mycelial form of Yarrowia
lipolytica.";
RL Yeast 12:1535-1548(1996).
DR EMBL; 281006; CAB02634.1; -.
KW Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 285 POTENTIAL.
SQ SEQUENCE 285 AA; 30036 MW; 0BA7FF1C515CD57D CRC64;

Query Match 13.2%; Score 109; DB 3; Length 285;
Best Local Similarity 28.4%; Pred. No. 0.012;
Matches 40; Conservative 11; Mismatches 62; Indels 28; Gaps 6;

OY 1 MCHSRCHPTMTILOAQTAP-STIPGPRRGSGPEITFDPLPEPAAAPAGRPSASRGHR 59
Db 122 ICHTKVTLSVTLPTTPVQTTTAVPEKPTPEVPEVKPEPTP-EVPGVKPEPTGPP 180
OY 60 KRRRVLYPRVVRQLP-VVEPNPAKRLLFLLLTIVFCQILMAEEGVPAPLPPEDAFNAA 118
Db 181 APKPEPEVPEVKPEPTPEVPEVRP-----EPTPAPLPPR--PSLR 218
OY 119 SLAPTPVSPVLEPFNLTSPTS 139
Db 219 SLRS---SPSLPLPLPPSPS 236

RESULT 11
O15047 PRELIMINARY; PRT; 1709 AA.
AC O15047;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE KIAA0339 PROTEIN (FRAGMENT).
GN KIAA0339
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res 4:141-150(1997).
DR EMBL; AB002337; BAA20797.2; -.
DR INTERPRO; IPR000504; -.
DR INTERPRO; IPR01214; -.
DR INTERPRO; IPR002965; -.
DR PFAM; PF000076; rrm; 1.
DR PFAM; PF00856; SET; 1.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
SQ SEQUENCE 1709 AA; 186231 MW; DCF2E0FF716B672C CRC64;

Query Match 13.2%; Score 109; DB 4; Length 1709;
Best Local Similarity 32.0%; Pred. No. 0.069;
Matches 44; Conservative 10; Mismatches 53; Indels 28; Gaps 7;

OY 6 SCIPTMTILOAQTAPSTIPGPRRGSGPEITFDPLPEPAAAPAGRPSASRGHRKRRV 65
Db 1078 SASPPPEVPVPTAPVPEVPPVPERVAGSPV---TPLPEQASPA-RPACHTESPSAPL 1133
OY 66 LYPRVVRRLPVEEPNPAKR-----LLFL-----LTIVFCQILMAEEGVPAPLPP 111
Db 1134 RPP-----EPPAGPAPAPRDERPSPSIPLLPPPKKRRKTVSFSAI----EVPVPEPP 1184
```

```
OY 112 EDAPNAASLAPTPVS 126
Db 1185 PATPPQAKF-PGPAS 1198

RESULT 12
Q91BT9 PRELIMINARY; PRT; 3325 AA.
ID Q91BT9;
AC Q91BT9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE MAJOR TEGUMENT PROTEIN.
GN UL36.
OS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
OX NCBI_TaxID=10390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GA;
RX MEDLINE=92237304; PubMed=1315048;
RA Jones D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;
RT "Marek disease virus encodes a basic-leucine zipper gene resembling
the fos/jun oncogenes that is highly expressed in lymphoblastoid
tumors.";
RT tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GA;
RA Lee L.F., Wu P., Sui D., Ren D., Kung H.J.;
RT "The Complete UL Sequence of Serotype I Marek's Disease Virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).
DR EMBL; AF147806; AAF66771.1; -.
SQ SEQUENCE 3325 AA; 365698 MW; 5D4A84EF719BB9FD CRC64;

Query Match 13.2%; Score 109; DB 12; Length 3325;
Best Local Similarity 25.2%; Pred. No. 0.13;
Matches 31; Conservative 14; Mismatches 52; Indels 26; Gaps 3;

OY 16 APTPAPSTIPGPRRGSGPEITFDPLPEPAAAPAGRPSASRGHRKRRVLYPRVVRQL 75
Db 2802 SPAPKSPAPKPKPPDPD---FKPSAPKPSAPKSPASPKSPKPPAPDSKPS 2858
OY 76 PVEEPNPAKRLLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASLAPTP-----VSPV 128
Db 2859 PAPKPKPPP-----TPDSKPSAPKPSASPKPLPVPPPNDSKSTSPV 2902

OY 129 LEP 131
Db 2903 PNP 2905

RESULT 13
Q9L568 PRELIMINARY; PRT; 233 AA.
ID Q9L568;
AC Q9L568;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PSPA (FRAGMENT).
GN PSPA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=39;
RA Beall B.W.;
RT "psa sequence types from multiresistant pneumococci.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
```

AC	Q9P944;	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DE	KEXIN-LIKE PROTEASE KEX1.	
CN	KEX1.	
OS	Pneumocystis carinii f. sp. muris.	
OC	Eukaryota; Fungi; Fungi incertae sedis; Pneumocystidaceae;	
OC	Pneumocystis.	
OX	NCBI_TaxID=42066;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Lee L.H., Gliotti F., Wright T.W., Simpson-Haidaris P.J.,	
RT	Weinberg G.A., Haidaris C.G.;	
RT	"Molecular characterization of KEX1, a kexin-like protease in mouse	
RT	Pneumocystis carinii.";	
RL	Gene 242:141-150(2000).	
DR	EMBL: AF093132; AAF32493.1;	
KW	Protease.	
QY	SEQUENCE 1011 AA; 112020 MW; FBE472C8F65864E8_CRC64;	
Query Match 13.4%; Score 110.5; DB 3; Length 1011;		
Best Local Similarity 25.0%; Pred. No. 0.03;		
Matches 40; Conservative 10; Mismatches 73; Indels 37; Gaps		
QY	4 SRSCPTMTLQAPTAPSTIPGRGSGPEIF-----TFDPLPEAPAAAGRPSASRGH 58	
Db		
QY	59 RKRSRRVLYPRVVRQLPVPEEPNPKRLLFLLLTI VFCOILMAEGVAPLPPEDA--- 114	
Db	722 EPTSEPTQPAPQPAPQPAPQP-----APQAPQPAPQPAPQP 763	
QY	115 -----PNAASLAPTPVSPVLEPNLTSEPSDYALDLST 147	
Db	764 VPPQVPVPPMPSPRPAPKPTQP---TSEPAPQTSEST 800	
RESULT 8		
QY6343	Q96343 PRELIMINARY; PRT; 552 AA.	
ID	Q96343	
OC	Q96343; 01-FEB-1997 (TrEMBLrel. 02, Created)	
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)	
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)	
DE	MYROSINASE-BINDING PROTEIN RELATED PROTEIN (FRAGMENT).	
OS	Brassica napus (Rape).	
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;	
OX	Brassicales; Brassicaceae; Brassica.	
OX	NCBI_TaxID=3706;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-20516 OF SVALOFS KARAT;	
RX	MEDLINE=97210758; Pubmed=9057822;	
RA	Taipalensuu J., Falk A., Ek B., Rask L.;	
RT	"Myrosinase-binding proteins are derived from a large wound-inducible	
RT	and repetitive transcript.";	
RL	Eur. J. Biochem. 243:605-611(1997).	
DR	EMBL: U59446; AAC08051.1;	
DR	HSSP; P18674; 1JOT.	
DR	MENDEL; 14865; Brana:2388;14865.	
DR	INTERPRO: IPR001229;	
DR	PFAM: PF01419; Jacalin; 3.	
FT	NON_TER 1	
QY	SEQUENCE 552 AA; 59270 MW; 5FE51A3718FBD0FA_CRC64;	
Query Match 13.3%; Score 109.5; DB 10; Length 552;		
Best Local Similarity 29.4%; Pred. No. 0.021;		
Matches 42; Conservative 6; Mismatches 42; Indels 53; Gaps		

Db 181 DLSTFLQHPAAF 193
|||||

RESULT 2

Q9NXH8 PRELIMINARY: PRT: 423 AA.
AC Q9NXH8;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE CDNA FLJ20245 FIS, CLONE COLF6454.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON MUCOSA;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isoigai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000252; BAA91032.1; -
SQ SEQUENCE 423 AA; 46942 MW; F3F52BC058CAD51E CRC64;

Query Match 14.9%; Score 122.5; DB 4; Length 423;
Best Local Similarity 42.2%; Pred. No. 0.001; 34; Indels 9; Gaps 5;
Matches 38; Conservative 9; Mismatches 34; Indels 9; Gaps 5;

QY 19 PAPSTIP-GPRRG-SQPEITFDPLPEPAAAPGRPSASRGHRKRRRLVPRVVRQLP 76
Db 58 PRPGCSRAPRADLPQKFTED---SPAELPSRTP---RKKRRSRRLVLYPETSRYKYP 111
QY 77 -VEEPNPAKRLFLLLTIVFCQILMAEGV 105
Db 112 RVEHRSRAQRCLLLLVIAIVGQVLNAIENL 141.

RESULT 3
O15026 PRELIMINARY: PRT: 1668 AA.
AC O15026;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE KIAA0309 (FRAGMENT).
GN KIAA0309.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002307; BAA20768.1; -
DR INTERPRO; IPR000637; -
DR INTERPRO; IPR001650; -
DR PFAM; PF00271; helicase_C; 1.
DR PRINTS; PR002178; AT_hook; 3.
DR NON_TER 1
SQ SEQUENCE 1668 AA; 175162 MW; 7717C8FC6A396C49 CRC64;

Query Match 13.9%; Score 114.5; DB 4; Length 1668;
Best Local Similarity 29.3%; Pred. No. 0.021; 50; Indels 37; Gaps 6;
Matches 41; Conservative 12; Mismatches 12; Indels 37; Gaps 6;

QY 6 SCHPTMTILOAP--TPAPSTIPGPRRGSGPEITFDPLPEPAAAPGRPSASRGHRKRSR 63
Db 854 SAHQTRSTTTPRCSPAREVRP-----APR-----PRTPASAPAAIPA----- 894
QY 64 RVLVPRVVRQLPVVEPNPAKRLFLLLTIVFCQILMAEGVPAPLPPEDAPNAASLAPT 123
Db 895 --LVPVPSAPVPISAPNP-----ITILPVHIL-----PSPPPPSQIPPCSSPACT 938
QY 124 PVSPVLEPFLNTSEPSDYAL 143
Db 939 PPPACTPPPAHTPPPAQTCL 958

RESULT 4
Q9Y5L9 PRELIMINARY: PRT: 2971 AA.
AC Q9Y5L9;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR SRCAP.
GN SRCAP.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99278407; PubMed=10347196;
RA Johnston H., Kneer J., Chackalaparampill I., Yaciuk P., Chriwia J.;
RT "Identification of a novel SNF2/SWI2 protein family member, SRCAP,
which interacts with CREB-binding protein."
RL J. Biol. Chem. 274:16370-16376(1999).
DR EMBL; AF143946; AAD39760.1; -
DR INTERPRO; IPR000330; -
DR INTERPRO; IPR000637; -
DR INTERPRO; IPR002965; -
DR INTERPRO; IPR001650; -
DR PFAM; PF00176; SNF2_N; 1.
DR PFAM; PF00271; helicase_C; 1.
DR PRINTS; PR00929; ATHOOK.
DR PRINTS; PR01217; PRICHXTENS.
SQ SEQUENCE 2971 AA; 315641 MW; F9F7EE70304B78A3 CRC64;

Query Match 13.9%; Score 114.5; DB 4; Length 2971;
Best Local Similarity 29.3%; Pred. No. 0.037;
Matches 41; Conservative 12; Mismatches 12; Indels 37; Gaps 6;

QY 6 SCHPTMTILOAP--TPAPSTIPGPRRGSGPEITFDPLPEPAAAPGRPSASRGHRKRSR 63
Db 2157 SAHQTRSTTTPRCSPAREVRP-----APR-----PRTPASAPAAIPA----- 2197
QY 64 RVLVPRVVRQLPVVEPNPAKRLFLLLTIVFCQILMAEGVPAPLPPEDAPNAASLAPT 123
Db 2198 --LVPVPSAPVPISAPNP-----ITILPVHIL-----PSPPPPSQIPPCSSPACT 2241
QY 124 PVSPVLEPFLNTSEPSDYAL 143
Db 2242 PPPACTPPPAHTPPPAQTCL 2261

RESULT 5
Q65553 PRELIMINARY: PRT: 3247 AA.
AC Q65553;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:04:55 ; Search time 64.84 Seconds
(without alignments)
281.993 Million cell updates/sec

Title: US-08-799-910-10
Perfect score: 823
Sequence: 1 MCHSRCHPTWTILOAPTPA.....EPSDYALDSTFLOQHPAAF 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL15.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	771.5	93.7	193	4	075353 homo sapien
2	122.5	14.9	423	4	09NXH8 homo sapien
3	114.5	13.9	1668	4	015026 homo sapien
4	114.5	13.9	2971	4	09Y519 homo sapien
5	114	13.9	3247	12	Q65553 bovine herp
6	112.5	13.7	649	5	09V884 drosophila
7	110.5	13.4	1011	3	09P944 pneumocysti
8	109.5	13.3	552	10	Q96343 brassica na
9	109.5	13.3	1315	10	Q9SPM0 zea mays (m
10	109	13.2	285	3	P78977 yarrowia li
11	109	13.2	1709	4	015047 homo sapien
12	109	13.2	3325	12	09IBT9 turkey herp
13	108.5	13.2	233	2	09L568 streptococc
14	107.5	13.1	203	2	09RCX9 streptomyc
15	107.5	13.1	581	5	Q20517 caenorhabdi
16	107.5	13.1	1013	5	09VYT7 drosophila
17	107	13.0	426	2	09LAY5 streptococc
18	107	13.0	496	12	Q98457 paramecium
19	106.5	12.9	901	2	Q44562 actinomyc

```

20 106.5 12.9 1498 4 09Y4F5
21 106 12.9 1029 12 09YMX0
22 106 12.9 2066 12 09YMX0
23 105.5 12.8 249 2 09L570
24 105 12.8 296 5 09V729
25 105 12.8 351 10 039492
26 105 12.8 549 12 089370
27 104.5 12.7 243 2 09L567
28 104.5 12.7 244 2 09L565
29 104.5 12.7 535 2 09L565
30 104.5 12.7 616 12 096716
31 104 12.6 395 2 09LAY2
32 104 12.6 408 2 09LAY0
33 103 12.5 241 5 045522
34 102.5 12.5 288 12 084565
35 102.5 12.5 767 5 062537
36 102 12.4 236 2 09L569
37 102 12.4 331 5 09V728
38 102 12.4 708 10 09SX31
39 101.5 12.3 344 11 089037
40 101 12.3 339 12 084465
41 101 12.3 395 2 09LAZ1
42 101 12.3 437 2 09LAY4
43 101 12.3 739 2 09ROT4
44 101 12.3 852 4 09NP71
45 100 12.2 802 11 070433

```

ALIGNMENTS

```

RESULT 1
ID 075353 PRELIMINARY; PRT; 193 AA.
AC 075353;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ANTI-DEATH PROTEIN.
GN IEX-1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98369175; PubMed=9703517;
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival.";
RL Science 281:998-1001(1998).
DR EMBL; AF039067; AAC32558.1; -.
DR ENBL; AF071596; AAC72344.1; -.
SQ SEQUENCE 193 AA; 21028 MW; 7927D9D3FFBC7C57 CRC64;

```

```

Query Match 93.7%; Score 771.5; DB 4; Length 193;
Best Local Similarity 78.8%; Pred. No. 8.1e-64;
Matches 152; Conservative 1; Mismatches 3; Indels 37; Gaps 1;

Qy 1 MCHSRCHPTWTILOAPTPAPSTIPGPRRGSGPEIFTFDPLPEAPAAAPAGPSASRGHRK 60
Db 1 MCHSRCHPTWTILOAPTPAPSTIPGPRRGSGPEIFTFDPLPEAPAAAPAGPSASRGHRK 60
Qy 61 RSRRLVYPR-----VVRRQLPVEENPA 83
Db 61 RSRRLVYPR-----VVRRQLPVEENPA 120
Qy 84 KRLFLLLTIVFCOILMAEGVAPLPPEADPNAAALAPTVPVSLGPNLTSPSDYAL 143
Db 121 KRLFLLLTIVFCOILMAEGVAPLPPEADPNAAALAPTVPVSLGPNLTSPSDYAL 180
Qy 144 DLSTFLQHPAAF 156

```

OY 21 LTIVFCQILMA-----EEGVPAFLPPEDAPNAASLAPTPVSPVLEPNTLSEP-SDYAL 73
db 793 ittvscgkmaaptrntctgttdpppgdgsnitsvshsvkvkfsqfeahgpkayav 852
OY 74 DLSTFLQOHPPAA 85
Db 853 ilttgeaghpaa 864

RESULT 15
W37148
ID W37148 standard: Protein: 541 AA.
XX AC W37148;
XX
XX 06-JUL-1998 (first entry)
XX Mammalian Ena (Mena).
XX Mena protein; mammalian Ena; Enabled; Evi protein; cytoskeleton;
KW cell morphology; cell adhesion; cell differentiation; cell growth;
KW cell motility; mouse.
XX
XX Mus musculus.

Key	Location/Qualifiers
Domain	1..113
Peptide	/note= "Ena-VASP homology domain 1 (EVH1)"
Peptide	156..160
Peptide	/note= "LERER repeat"
Peptide	166..170
Peptide	/note= "LERER repeat"
Peptide	171..175
Peptide	/note= "LERER repeat"
Peptide	176..180
Peptide	/note= "LERER repeat"
Peptide	199..203
Peptide	/note= "LERER repeat"
Peptide	207..211
Peptide	/note= "LERER repeat"
Modified-site	236
Binding-site	/note= "O-phosphorylated"
Binding-site	305..310
Binding-site	/note= "mediates binding to profilin"
Binding-site	316..321
Binding-site	/note= "mediates binding to profilin"

WO9801755-A1.
XX
XX 15-JAN-1998.
XX
XX 03-JUL-1997; 97WO-US11669.
XX
XX 05-JUL-1996; 96US-0675815.
XX
XX (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX Certlier FB, Niebuhr K, Soriano P, Wehland J;
XX
XX WPI: 1998-101197/09.
XX N-PSDB; V02996.
XX
XX Detection of modulators of Mena and Ena-VASP-like genes and proteins
XX - used in control of cytoskeletal dynamic events in normal and
XX abnormal cell morphology, adhesion, motility, growth and
XX differentiation
XX
XX Example 1: Page 54-56; 77pp; English.
XX
XX This polypeptide comprises murine Mena (mammalian Ena) that shows
XX significant amino acid similarity to Drosophila Ena and which
XX exhibits a broad pattern of tissue distribution in neurons,

CC fibroblasts, kidney epithelium, muscle, neural crest and
CC haematopoietic cells. Its amino acid sequence was deduced from
CC a cDNA clone (see V02996) obtained from a mouse embryonic stem cell
CC cDNA library. 2 Novel mammalian genes, Mena and Ena-VASP-like
CC (Evi, see V02997), encoding novel proteins Mena and Evi (see
CC W37148) are disclosed. Mena and Evi proteins have a discrete, EVH1
CC functional domain responsible for Mena binding to Listeria, and to
CC the cytoskeletal proteins zyxin and vinculin. The EVH1 domain of
CC Mena is also responsible and sufficient for targeting localisation
CC of Mena and Mena-based fusion proteins to focal adhesions, and to
CC the surface of Listeria cells at the polar site of induction of
CC actin comet tail formation. Based on the disclosed Mena and Evi
CC genes and proteins, a variety of methods and compositions are
CC provided for screening, isolating and characterising endogenous and
CC exogenous factors, drugs and therapeutic agents useful to evaluate
CC and/or control cytoskeletal dynamic events involved in normal and
CC abnormal cell morphology, adhesion, motility, growth and/or
CC differentiation. A method of detecting a modulator of Mena
CC activity/expression is claimed.
XX
XX Sequence 541 AA;
SQ

Query Match 15.28; Score 67; DB 19; Length 541;
Best Local Similarity 29.9%; Pred. No. 25;
Matches 20; Conservative 4; Mismatches 23; Indels 20; Gaps 2;
OY 6 PVPEPNPAKRLLLFLLTIVFCQILMAEEGVPAFLPPEDAPN-----AASLAPTPVSPVLEP 61
Db 268 paesptpqgivil-----gppappppplpssgpayasalppppgppppp 311
OY 62 FNLTSEP 68
Db 312 lpstgpp 318

Search completed: January 30, 2001, 22:05:37
Job time: 11886 sec

DT 08-AUG-2000 (first entry)
XX Human PRO1431 (UNQ737) amino acid sequence SEQ ID NO:315.
DE
XX
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS Homo sapiens.
XX WO200012708-A2.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0098536.
PR 09-SEP-1998; 98US-0098596.
PR 09-SEP-1998; 98US-0098598.
PR 09-SEP-1998; 98US-0098602.
PR 09-SEP-1998; 98US-0098642.
PR 10-SEP-1998; 98US-0098741.
PR 10-SEP-1998; 98US-0098754.
PR 10-SEP-1998; 98US-0098763.
PR 10-SEP-1998; 98US-0098792.
PR 10-SEP-1998; 98US-0098808.
PR 10-SEP-1998; 98US-0098812.
PR 10-SEP-1998; 98US-0098815.
PR 10-SEP-1998; 98US-0098816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100384.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 24-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.

PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 30-OCT-1998; 98US-0106500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
PI WPI: 2000-237871/20.
XX N-PSDB: A37110.
DR
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
PT


```

RESULT 6
R89436
ID R89436 standard; Protein; 610 AA.
XX AC
XX R89436;
XX
XX 02-SEP-1996 (first entry)
XX DE
XX Mutated platelet glycoprotein-Ib-alpha GPIb protein sequence.
XX KW
XX Platelet glycoprotein-Ib-alpha; GPIb; mutagenesis; point mutation;
XX KW
XX von Willebrand factor; blood disorder; platelet disorder;
XX KW
XX protein engineering; Bernard-Soulier disease.
XX OS
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 36...200
XX FT /note= "substitution from Phe in wild-typ
XX FT /note= "GPIb-alpha"
XX FT 200...220
XX FT /note= "Leu rich area"
XX FT 220...310
XX FT /note= "flanking region to Leu rich area"
XX FT 310...420
XX FT /note= "hinge region"
XX FT /note= "Ser/Thr rich area"
XX
XX US5492809-A.
XX PN
XX
XX 20-FEB-1996.
XX PD
XX
XX 07-OCT-1991; 91US-0770968.
XX PF
XX
XX 15-JAN-1992; 92US-0821717.
XX PR
XX 07-OCT-1991; 91US-0770968.
XX PR
XX 09-SEP-1993; 93US-0119262.
XX PR
XX (UYNV ) UNIV NEW YORK STATE RES FOUND.
XX PA
XX Cunningham D, Finch CN, Lyle VA, Miller JL;
XX PI
XX WPI; 1996-128585/13.
XX DR
XX
XX DNA encoding platelet glyco:protein Ib alpha mutant Phe57 -
XX PT
XX introduced into platelets to reduce aggregation and reactivity with
XX PT
XX von Willebrand factor; also probe for diagnosis of Bernard-Soulier
XX PT
XX disease
XX PS
XX Disclosure; Column 21-26; 20pp; English.
XX
XX A substitution of T for C at position 259 in the DNA sequence of
XX CC
XX GPIb-alpha leads to the replacement of Phe for Leu at
XX CC
XX residue 57 of the mature GPIb-alpha molecule. This mutated GPIb-
XX CC
XX alpha protein is less reactive with von Willebrand factor, a
XX CC
XX characteristic of the autosomal recessive bleeding disorder,
XX CC
XX Bernard-Soulier disease.
XX
XX Sequence 610 AA:
XX
Query Match 16.2%; Score 71.5; DB 17; Length 610;
Best Local Similarity 45.08; Pred. NO. 9.5;
Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;
QY 36 PAPLP-----PEDAPNAASLAPTPSPVLEPFLNLTSEPS 69
Db 366 ptpsttsepyepapmmttleptpspttpep---tsepa 402

```

RESULT 7

```

W18201
ID W18201 standard; protein; 610 AA.
XX AC
XX W18201;
XX
XX 18-AUG-1997 (first entry)
XX DE
XX Platelet glycoprotein Ib alpha naturally-occurring wild-type.
XX KW
XX GPIb alpha; recombinant polypeptide; antithrombotic;
XX KW
XX platelet adhesion; platelet aggregation; thrombus formation;
XX KW
XX platelet-type von Willebrand disease.
XX OS
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 36...200
XX FT /label= Leucine-rich
XX FT 220...310
XX FT /label= Hinge
XX FT /note= "A major binding site for von Willebrand factor"
XX FT 228...238
XX FT /note= "Preferred sites for mutations that result
XX FT in a mutant polypeptide having more reactivity
XX FT with von Willebrand factor"
XX
XX Misc-difference 233
XX FT /note= "Preferred mutation; substitution of Val for Gly"
XX FT 310...420
XX FT /label= Serine/threonine-rich
XX
XX US5624817-A.
XX PN
XX
XX 29-APR-1997.
XX PD
XX
XX 07-OCT-1991; 91US-0770968.
XX PF
XX
XX 07-OCT-1991; 91US-0770968.
XX PR
XX 28-APR-1994; 94US-0234265.
XX PR
XX (UYNV ) UNIV NEW YORK STATE RES FOUND.
XX PA
XX Cunningham D, Finch CN, Lyle VA, Miller JL, Pincus MR;
XX PI
XX WPI; 1997-258215/23.
XX DR
XX
XX DNA encoding mutant platelet glyco:protein Ib alpha - for production
XX PT
XX of recombinant polypeptide(s) useful as antithrombotic agents, etc.
XX PS
XX Disclosure; Column 27-32; 20pp; English.
XX
XX The present sequence represents the naturally-occurring wild-type
XX CC
XX platelet glycoprotein Ib alpha (GPIb alpha). The polypeptide can have a
XX CC
XX mutation which makes it more reactive with von Willebrand factor (vWF)
XX CC
XX than the wild-type. Preferably the mutation occurs in the hinge region,
XX CC
XX within residues 228 to 238. The preferred mutation is a substitution of
XX CC
XX Val for the wild-type Gly at position 233. The mutation alters the
XX CC
XX three-dimensional structure of the mutant polypeptide from a beta-bend
XX CC
XX conformation to an alpha helix conformation, and also creates an
XX CC
XX amphipathic region within the polypeptide. The DNA encoding mutant
XX CC
XX platelet GPIb alpha polypeptides (where the mutation is between the
XX CC
XX residues 228 to 238, but not at position 233), vectors and cells are
XX CC
XX useful for production of recombinant mutant GPIb alpha polypeptides,
XX CC
XX which can be used to inhibit platelet adhesion and aggregation (e.g. in
XX CC
XX antithrombotic medications or on medical devices or vascular
XX CC
XX prostheses), or can be labelled and used as imaging agents, or can be
XX CC
XX coupled to a thrombolytic agent for targeting to sites of thrombus
XX CC
XX formation. The oligonucleotides can be used as probes for diagnosis of
XX CC
XX platelet-type von Willebrand disease.
XX
XX Sequence 610 AA:
XX
Query Match 16.2%; Score 71.5; DB 18; Length 610;

```

CC three-dimensional coordinates obtained from the crystals, is useful
 CC for identifying an agent that stabilizes the Ras-Sos complex. The
 CC crystals are also useful for identifying agents that inhibit the
 CC formation of Ras-Sos complex. Ras and Sos fragments are useful for
 CC growing a crystal of a protein-ligand complex. Agents that stabilize
 CC or inhibit the formation of Ras-Sos complex are useful in the
 CC treatment of cancer. The present sequence represents a Drosophila
 CC Sos protien.

XX SQ Sequence 1596 AA;

Query Match 18.0%; Score 79.5; DB 21; Length 1596;
 Best Local Similarity 28.3%; Pred. No. 4.2;
 Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

OY 1 VRRQLPVE-----EPNPAKRLLLFTVFCQILMAEGVPAPLPEDAPNAASLAPT 53
 Db 1423 irnsalekraaatsqnaagpisttlvtvsqavatdeplplisp-----aasstt 1477
 OY 54 --PVSPVLEPF--NLTSEP-----SDYALDLSTFLOQ-----HPAAF 86
 Db 1478 tspltpamspnlpshpvestsssyahqlrmrqqqqqqthpaly 1523

RESULT 4

RS1116 ID R51116 standard; protein; 610 AA.

XX AC R51116;

DT 23-SEP-1994 (first entry)

XX DE Platelet glycoprotein Ib alpha.

XX KW Platelet; glycoprotein; imaging; thrombolytic agent;
 KW tissue plasminogen activator; tPA; pro-urokinase; urokinase;
 KW streptokinase; Bernard-Soulier disease; thrombus; aggregation;
 KW anisoylated plasminogen-streptokinase activator complex; adhesion;
 KW inhibition.

XX OS Homo sapiens.

XX PN US5298239-A.

XX PD 29-MAR-1994.

XX PF 07-OCT-1991; 91US-0770968.

XX PR 07-OCT-1991; 91US-0770968.

XX PR 15-JAN-1992; 92US-0821717.

XX PA (UUNY) UNIV NEW YORK STATE RES FOUND.

XX PI Cunningham D, Finch CN, Lyle VA, Miller JL;

XX DR WPI; 1994-100287/12.

XX PT Platelet glycoprotein Ib alpha with an amino acid substn at
 PT position 57 - has reduced reactivity with Von Willebrand factor,
 PT and can be used to inhibit platelet aggregation and inhibition

XX PS Claim 1; Columns 3-8; 20pp; English.

XX CC A substitution in platelet glycoprotein Ib alpha (Leucine 57 to
 CC Phenylalanine) underlies a form of Bernard-Soulier disease. The
 CC mutated glycoprotein can be used in compositions to inhibit
 CC platelet aggregation/adhesion. The glycoprotein may be labelled and
 CC used as an imaging agent and may also be bound to a thrombolytic
 CC agent, preferably tissue plasminogen activator (tPA),
 CC (pro) urokinase, streptokinase, anisoylated plasminogen-streptokinase
 CC activator complex, tPA analogues or a protease, allowing localisation
 CC of the thrombolytic agent to a thrombus.

XX SQ Sequence 610 AA;

Query Match 16.2%; Score 71.5; DB 15; Length 610;
 Best Local Similarity 45.0%; Pred. No. 9.5;
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

OY 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFPNLTSEPS 69
 Db 366 ptpsttsepyepapapmttleptpspttpep----tsepa 402

RESULT 5

RS5664 ID R5664 standard; protein; 610 AA.

XX AC R5664;

DT 23-SEP-1994 (first entry)

XX DE Mutant platelet glycoprotein Ib alpha.

XX KW Platelet; glycoprotein; imaging; thrombolytic agent;
 KW tissue plasminogen activator; tPA; pro-urokinase; urokinase;
 KW streptokinase; Bernard-Soulier disease; thrombus; aggregation;
 KW anisoylated plasminogen-streptokinase activator complex; adhesion;
 KW inhibition.

XX OS Homo sapiens.

XX PN US5298239-A.

XX PD 29-MAR-1994.

XX PF 07-OCT-1991; 91US-0770968.

XX PR 07-OCT-1991; 91US-0770968.

XX PR 15-JAN-1992; 92US-0821717.

XX PA (UUNY) UNIV NEW YORK STATE RES FOUND.

XX PI Cunningham D, Finch CN, Lyle VA, Miller JL;

XX DR WPI; 1994-100287/12.

XX PT Platelet glycoprotein Ib alpha with an amino acid substn at
 PT position 57 - has reduced reactivity with Von Willebrand factor,
 PT and can be used to inhibit platelet aggregation and inhibition

XX PS Claim 1; 20pp; English.

XX CC A substitution in platelet glycoprotein Ib alpha (Leucine 57 to
 CC Phenylalanine) underlies a form of Bernard-Soulier disease. The
 CC mutated glycoprotein can be used in compositions to inhibit
 CC platelet aggregation/adhesion. The glycoprotein may be labelled and
 CC used as an imaging agent and may also be bound to a thrombolytic
 CC agent, preferably tissue plasminogen activator (tPA),
 CC (pro) urokinase, streptokinase, anisoylated plasminogen-streptokinase
 CC activator complex, tPA analogues or a protease, allowing localisation
 CC of the thrombolytic agent to a thrombus.

XX SQ Sequence 610 AA;

Query Match 16.2%; Score 71.5; DB 15; Length 610;
 Best Local Similarity 45.0%; Pred. No. 9.5;
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

OY 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFPNLTSEPS 69
 Db 366 ptpsttsepyepapapmttleptpspttpep----tsepa 402

PT disease, e.g. atherosclerosis, restenosis, hypertension, etc
 XX
 PS Example 6; Fig 5; 163pp; English.

CC This protein is encoded by the novel human fchd605 gene (see
 CC T94471) that is up-regulated in monocytes treated with oxidised
 CC low density lipoproteins that simulate the conditions under which
 CC foam cells develop during atherogenesis. The protein has sequence
 CC similarity to the mouse gly96 gene and to EST T49532. Novel
 CC fchd531, fchd540, fchd545, fchd602 and fchd605 genes (see T94467-71)
 CC provide a fingerprint for the study of cardiovascular diseases,
 CC including atherosclerosis, ischaemia/reperfusion, hypertension,
 CC restenosis and arterial inflammation. Methods are provided for the
 CC diagnosis, monitoring in clinical trials, screening for
 CC therapeutically effective compounds, and treatment of
 CC cardiovascular diseases based on discoveries regarding the
 CC expression patterns of these novel genes.

XX Sequence 156 AA:

Query Match 100.0%; Score 442; DB 18; Length 156;
 Best Local Similarity 100.0%; Pred. No. 1.5e-39;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VRRQLPVEPNPAKRLLLFTIVFCQILMAEEGVPAPLPDPADAPNAASLAPTVPSPVLE 60
 DB 71 VRRQLPVEPNPAKRLLLFTIVFCQILMAEEGVPAPLPDPADAPNAASLAPTVPSPVLE 130

OY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86
 DB 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 2

Y45017
 ID Y45017 standard; Protein: 156 AA.

XX Y45017:

XX 31-MAY-2000 (first entry)

DE Protein encoded by fchd605 gene.

KW fchd605 gene; human; cardiovascular disease; oncogenic disorder;
 KW diabetic retinopathy; fibroproliferative disorder; atherosclerosis;
 KW TGF-beta signalling pathway; TGF-beta transforming growth factor;
 KW pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;
 KW vascularisation; cytostatic; antidiabetic; ophthalmological.

XX Homo sapiens.

XX WO200006206-A1.

XX 10-FEB-2000.

XX 30-JUL-1999; 99WO-US17394.

XX 30-JUL-1998; 98US-0126640.

XX (MILL-) MILLENNIUM PHARM INC.

XX Falb DA:

XX WPI: 2000-205414/18.

XX N-PSDB; 250711.

XX Identifying substances for ameliorating symptoms of fibroproliferative
 XX diseases or oncogenic related disorders.

XX Example; Fig 5; 214pp; English.

XX The patent discloses methods for the treatment and diagnosis of

CC cardiovascular diseases by novel human genes which are differentially
 CC expressed in different cardiovascular disease states. Compositions which
 CC can modify TGF-beta signalling pathway are identified by screening.
 CC These are used therapeutically to treat fibroproliferative and oncogenic
 CC disorders, especially tcf (transforming growth factor)-beta related
 CC disorders, including diabetic retinopathy, atherosclerosis, pancreatic
 CC cancer, angiogenesis, inflammation, fibrosis, tumour growth and
 CC vascularisation. The present sequence is the protein product of fchd605
 CC gene which is up-regulated in monocytes treated with oxidised LDL (low
 CC density lipoprotein) can be used to design cardiovascular disease
 CC treatment strategies. Depending on whether the up-regulation has a
 CC pathogenic or protective effect treatment methods can be designed to
 CC increase or decrease the activity of the protein product of the gene.

XX Sequence 156 AA:

Query Match 100.0%; Score 442; DB 21; Length 156;
 Best Local Similarity 100.0%; Pred. No. 1.5e-39;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VRRQLPVEPNPAKRLLLFTIVFCQILMAEEGVPAPLPDPADAPNAASLAPTVPSPVLE 60
 DB 71 VRRQLPVEPNPAKRLLLFTIVFCQILMAEEGVPAPLPDPADAPNAASLAPTVPSPVLE 130

OY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86
 DB 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 3

Y68821
 ID Y68821 standard; protein; 1596 AA.

XX Y68821:

XX 16-MAY-2000 (first entry)

XX Amino acid sequence of a Drosophila Son of sevenless (Sos) protein.

KW Ras; Son of sevenless; Sos; crystal; Ras-Sos complex; cancer;
 KW protein coordinate data.

XX Drosophila melanogaster.

XX WO200005258-A1.

XX 03-FEB-2000.

XX 20-JUL-1999; 99WO-US16348.

XX 21-JUL-1998; 98US-0119794.

XX (UYRQ) UNIV ROCKEFELLER.

XX Borlack-Sjodin A, Margarit SM, Bar-Sagi D, Cole P, Kuriyan J;

XX WPI: 2000-182647/16.

XX Novel crystals comprising a Ras-Son of sevenless complex, useful for
 XX screening drugs useful in cancer treatment

XX Disclosure; Page 169-175; 224pp; English.

XX The specification describes a crystal complex comprising at least a
 CC Sos contacting region of a Ras protein and at least a Son of sevenless
 CC (Sos) protein catalytic region fragment, that effectively diffracts
 CC x-rays. Ras and Sos form a tight complex. Sos does not impede the
 CC binding sites for the nucleotide base and the ribose of GTP or GDP
 CC and thus the Ras-Sos complex maintains a structure that permits
 CC nucleotide release and rebinding. The crystals are used for the
 CC determination of the atomic coordinates of the complex to a resolution
 CC of more than 5.0 Angstrom. The crystals, or a dataset comprising the

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:05:37 ; Search time 54.97 Seconds
(without alignments)
53.496 Million cell updates/sec

Title: US-08-799-910-10_COPY_71_156
Perfect score: 442
Sequence: 1 VRRQLPVEEPNPAKRLLEL.....EPDYALDLSTFLQHPAAF 86

Scoring table:
BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2_2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /cgn2_2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /cgn2_2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /cgn2_2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /cgn2_2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /cgn2_2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /cgn2_2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /cgn2_2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /cgn2_2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /cgn2_2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /cgn2_2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /cgn2_2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /cgn2_2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /cgn2_2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /cgn2_2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /cgn2_2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /cgn2_2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	442	100.0	156	18	W36006 Human Fchd605 gene
2	442	100.0	156	21	Y45017 Protein encoded by
3	79.5	18.0	1596	21	Y68821 Amino acid sequenc
4	71.5	16.2	610	15	R31116 Platelet glycoprot
5	71.5	16.2	610	15	R36664 Mutant platelet g
6	71.5	16.2	610	17	R89436 Mutated platelet g
7	71.5	16.2	610	18	W18201 platelet glycoprot
8	69.5	15.7	739	18	W25790 Gene 036 product d
9	68.5	15.5	204	18	W14578 Streptococcus pneu
10	68.5	15.5	370	21	Y99428 Human PRO1431 (UNQ
11	68.5	15.5	878	15	R55060 Sequence of human
12	68.5	15.5	878	16	R85487 Human E-cadherin p

13	68.5	15.5	882	20	Y09375 Wild-type E-cadher
14	68.5	15.5	1337	16	R85203 huDEP-1. Homo sap
15	67	15.2	541	19	Mammalian Ena (Men
16	67	15.2	783	19	Mouse neural Mena+
17	67	15.2	787	19	W37151 Mouse neural Mena+
18	67	15.2	802	19	W37152 Mouse neural Mena+
19	66.5	15.0	131	18	W27646 Secreted protein A
20	66.5	15.0	131	18	W4082 Human secreted pro
21	65.5	14.8	128	18	W14577 Streptococcus pneu
22	65.5	14.8	286	17	W03566 Mycobacterium tube
23	65.5	14.8	325	17	W03565 Mycobacterium tube
24	65.5	14.8	332	18	W32418 Mycobacterium tube
25	65.5	14.8	332	18	W32350 Mycobacterium tube
26	65.5	14.8	332	19	W81683 M. tuberculosis im
27	65.5	14.8	332	19	W64322 Mycobacterium tube
28	65.5	14.8	332	20	Y39083 M. tuberculosis an
29	65.5	14.8	332	20	Y38945 M. tuberculosis re
30	65.5	14.8	456	20	Y17067 Human 3-OST-4 prot
31	65.5	14.8	552	20	Y39225 M. tuberculosis fu
32	65.5	14.8	552	20	Y39082 M. tuberculosis fu
33	65.5	14.8	802	19	W81746 M. tuberculosis fu
34	65.5	14.8	802	19	W64379 Mycobacterium anti
35	65.5	14.8	802	20	Y32063 Mycobacterium tube
36	65.5	14.8	802	20	Y39224 M. tuberculosis fu
37	65.5	14.8	802	20	Y39176 M. tuberculosis fu
38	65.5	14.8	802	20	Y39081 M. tuberculosis fu
39	65.5	14.8	802	20	Y39033 M. tuberculosis fu
40	65.5	14.8	1093	16	R66460 AF-17 protein. Ho
41	65	14.7	328	19	W36984 E. coli ZipA prote
42	65	14.7	427	21	Y68748 Amino acid sequenc
43	64.5	14.6	147	15	R60799 Rape abscission/de
44	64.5	14.6	194	18	W14584 Streptococcus pneu
45	64.5	14.6	412	20	Y49151 Amino acid sequenc

ALIGNMENTS

RESULT	1
W36006	
ID	W36006 standard; Protein: 156 AA.
XX	
AC	W36006;
XX	
DT	03-MAR-1998 (first entry)
XX	
DE	Human Fchd605 gene product.
XX	
KW	Fchd605 gene; differential expression; monocyte; human;
KW	foam cell; cardiovascular disease; atherosclerosis; ischaemia;
KW	reperfusion; hypertension; restenosis; arterial inflammation;
KW	therapy; diagnosis; drug screening; marker.
OS	Homo sapiens.
XX	
PN	W09730065-A1.
XX	
PD	21-AUG-1997.
XX	
PF	14-FEB-1997; 97WO-US02291.
XX	
PR	13-FEB-1997; 97US-0799910.
PR	16-FEB-1996; 96US-0011787.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Falb DA;
XX	
DR	WPI; 1997-424966/39.
XX	
DR	N-PSDB; T94471.
XX	
PT	New genes differentially expressed in cardiovascular disease - used for diagnosis, drug screening and treatment of cardiovascular

Search completed: January 30, 2001, 22:06:26
Job time: 9705 sec

QY 29 LMAEEGVAPLPPEDAP---NAASLAPTVPSP 57
: : | | | | | | | : | | | | : |
Db 247 LVAPPPAPAPAAEPAPAPAPAPAGEVAPTPTTP 278

RESULT 15
 US-08-641-356-3
 ; Sequence 3, Application US/08641356
 ; Patent No. 5866130
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: MICROBACTERIAL PROTEINS,
 ; TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND THEIR USE FOR VACCINES
 ; TITLE OF INVENTION: AND FOR THE DETECTION OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER &
 ; ADDRESSEE: NEUSTADT, P.C.
 ; STREET: 1755 S. Jefferson Davis Highway, Suite 400
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/641,356
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/382,184
 ; FILING DATE: 01-FEB-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OBLON, NORMAN F.
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 22640720
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-413-3000
 ; TELEFAX: 703-413-2220
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 286 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; US-08-641-356-3

Matches 26: Conservative 11; Mismatches 31; Indels 33; Gaps 5;

QY 1 VRRQLVPEPNPAKRLLELLLTIVFC-----QILMAEEGVAPLPPEDAPNAASLA--- 51
DB 584 VNDNAPIPEPR-----TIFFCERNRPQVINIHD---ADLPPTSPTAELTHGR 630
QY 52 -----PTVPSPVLEPFLNTSEPSDYALDLSTFLQOH 82
DB 631 VPNWTIOYNDPTQESIILKP-KMALEVGDYKINKLMDNQN 670

RESULT 12
US-08-854-585-2
; Sequence 2, Application US/08854585
; Patent No. 611410
; GENERAL INFORMATION:
; APPLICANT: Tonks, Nicholas K. and stman, Arne
; TITLE OF INVENTION: Density Enhanced Protein Tyrosine Phosphatase
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, Suite 6300
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,585
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/237,940
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25,447
; REFERENCE/DOCKET NUMBER: 27866/31954
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 1337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-854-585-2

Query Match 15.5%; Score 68.5; DB 3; Length 1337;
Best Local Similarity 29.2%; Pred. No. 15;
Matches 21; Conservative 11; Mismatches 33; Indels 7; Gaps 2;

QY 21 LTIIVFCQILMA-----EGVAPLPPEDAPNAASLAPTPSPVLEPFLNTSEP-SDYAL 73
DB 793 ITTVSCGKMAAPTRNTCTTGITDPPDPGSPNITSVSHNSVKVFKSGFEASHGPIKAYAV 852
QY 74 DLSTFLQOHPPAA 85
DB 853 ILTTGEAGHPSA 864

RESULT 13
PCT-US95-05512-2
; Sequence 2, Application PC/TUS9505512
; GENERAL INFORMATION:
; APPLICANT: Tonks, Nicholas K. and stman, Arne
; TITLE OF INVENTION: Density Enhanced Protein Tyrosine
; Phosphatase

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 233 South Wacker Drive, Suite 6300
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05512
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27866/31954
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05512-2

Query Match 15.5%; Score 68.5; DB 4; Length 1337;
Best Local Similarity 29.2%; Pred. No. 15;
Matches 21; Conservative 11; Mismatches 33; Indels 7; Gaps 2;

QY 21 LTIIVFCQILMA-----EGVAPLPPEDAPNAASLAPTPSPVLEPFLNTSEP-SDYAL 73
DB 793 ITTVSCGKMAAPTRNTCTTGITDPPDPGSPNITSVSHNSVKVFKSGFEASHGPIKAYAV 852
QY 74 DLSTFLQOHPPAA 85
DB 853 ILTTGEAGHPSA 864

RESULT 14
US-08-382-184-3
; Sequence 3, Application US/08382184
; Patent No. 5714593
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MICROBACTERIAL PROTEINS,
; TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND THEIR USE FOR VACCINE
; TITLE OF INVENTION: AND FOR THE DETECTION OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER &
; ADDRESSEE: NEUSTADT, P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/382,184
; FILING DATE: 01-FEB-1995
; CLASSIFICATION: 435

```
;
;
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ For Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/035,648
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/818,829
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 739 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-035-648-24

Query Match 15.7%; Score 69.5; DB 3; Length 739;
Best Local Similarity 37.8%; Pred. No. 5.6;
Matches 17; Conservative 8; Mismatches 19; Indels 1; Gaps 1;

QY 25 FCOILMAEGVPAPLPEDAPNAASLAP-TPVSPVLEPFLNLTSEP 68
| | | | | | | | | | | | | | | | | | | | | |
DB 391 FMOSSSAKQTPPPVAPKPAVKSSSSQPTVPSVWSPGVAPTOP 435

RESULT 10
US-08-237-919-2
; Sequence 2, Application US/08237919
; Patent No. 5610281
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B
; APPLICANT: Cepek, Karyn L
; TITLE OF INVENTION: Methods and Compositions for
; TITLE OF INVENTION: Modulating Heterotypic E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/237,919
; FILING DATE:
; CLASSIFICATION: 424
```

```
;
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-237-919-2

Query Match 15.5%; Score 68.5; DB 1; Length 878;
Best Local Similarity 25.7%; Pred. No. 9;
Matches 26; Conservative 11; Mismatches 31; Indels 33; Gaps 5;

QY 1 VRRQLPVEPNPAKRLFLLLTIVFC-----QILMAEGVPAPLPEDAPNAASLA--- 51
| | | | | | | | | | | | | | | | | | | | | |
DB 584 VNDNAPIEPR-----TIFFCERNPKQVINHD---ADLPNTSPFTAE LTHGR 630

QY 52 -----PTPVSPLVLEPFLNLTSEPSDYALDLSTFLQOH 82
| | | | | | | | | | | | | | | | | | | | | |
DB 631 VPNTIYNDPTQESIILKP-KMALEVGDYKINLKLMNDNON 670

RESULT 11
PCT-US95-05518-2
; Sequence 2, Application PC/TUS9505518
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; TITLE OF INVENTION: E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05518
; FILING DATE: herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/237,919
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-05518-2

Query Match 15.5%; Score 68.5; DB 4; Length 878;
Best Local Similarity 25.7%; Pred. No. 9;
```

DATE: AUG-1987
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 610
PUBLICATION INFORMATION:
AUTHORS: Zimmerman, Theodore S.
AUTHORS: Ruggeri, Zaverio M.
AUTHORS: Houghten, Richard A.
AUTHORS: Vincete, Vincete
AUTHORS: Mohri, Hiroshi
TITLE: Proteolytic fragments and synthetic peptides
TITLE: that block the binding of von Willebrand
TITLE: membrane glycoprotein Ib
DOCUMENT NUMBER: EP 0 317 278 A2
FILING DATE: 16-NOV-1988
PUBLICATION DATE: 24-MAY-1989
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293
US-08-119-262B-6

Query Match 16.2% Score 71.5; DB 1; Length 610;
Best Local Similarity 45.0%; Pred. No. 2.7;
Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

Qy 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFLNLTSEPS 69
Db 366 PTPSPTTSEPVPEPAPNMTTLEPTSPPTPEP---TSEPA 402

RESULT 7
US-08-135-929A-11
Sequence 11, Application US/08135929A
Patent No. 5593959
GENERAL INFORMATION:
APPLICANT: Miller, Jonathan L.
APPLICANT: Cunningham, David
APPLICANT: Lyle, Vicki A.
APPLICANT: Finch, Clara N.
APPLICANT: Pincus, Matthew R.
TITLE OF INVENTION: Mutations in the Gene Encoding the Alpha
TITLE OF INVENTION: Chain of Platelet Glycoprotein Ib
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/135.929A
FILING DATE: 14-OCT-1993
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20884/23
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
TELEX: 978450

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-135-929A-11

factor to the

Query Match 16.2% Score 71.5; DB 1; Length 610;
Best Local Similarity 45.0%; Pred. No. 2.7;
Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

Qy 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFLNLTSEPS 69
Db 366 PTPSPTTSEPVPEPAPNMTTLEPTSPPTPEP---TSEPA 402

RESULT 8
US-08-234-265A-11
Sequence 11, Application US/08234265A
Patent No. 5624817

GENERAL INFORMATION:
APPLICANT: Miller, Jonathan L.
APPLICANT: Cunningham, David
APPLICANT: Lyle, Vicki A.
APPLICANT: Finch, Clara N.
APPLICANT: Pincus, Matthew R.
TITLE OF INVENTION: Mutations in the Gene Encoding the
TITLE OF INVENTION: Chain of Platelet Glycoprotein Ib
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/234.265A
FILING DATE: 28-APR-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20884/24
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
TELEX: 978450
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-234-265A-11

Query Match 16.2% Score 71.5; DB 1; Length 610;
Best Local Similarity 45.0%; Pred. No. 2.7;
Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

Qy 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFLNLTSEPS 69
Db 366 PTPSPTTSEPVPEPAPNMTTLEPTSPPTPEP---TSEPA 402

RESULT 9
US-09-035-648-24
Sequence 24, Application US/09035648
Patent No. 6100031

GENERAL INFORMATION:
APPLICANT: Shyjan, Andrew W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL
TITLE OF INVENTION: GROWTH AND PROLIFERATION

; FILING DATE: 16-NOV-1988
 ; PUBLICATION DATE: 24-MAY-1989
 ; PUBLICATION RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293
 US-07-821-717B-6

```

Query Match          16.2%; Score 71.5; DB 1; Length 610
Best Local Similarity 45.0%; Pred. No. 2.7;
Matches 18; Conservative 2; Mismatches 11; Indels

QY 36 PAPLP-----PEDAPNAAASLAPTPVSPVLPFNLITSEPS 69
      ||| ||| ||| ||| ||| ||| ||| |||
DB 366 PTPSPTTSEPVPEPAPNNTLTLEPTPSPTTPEP---TSEPA 402

RESULT 6
US-08-119-262B-6
: Sequence 6, Application US/081119262B
: Patent No. 5492809
: GENERAL INFORMATION:
: APPLICANT: Miller, Jonathan L.
: APPLICANT: Cunningham, David
: APPLICANT: Lyle, Vicki A.
: APPLICANT: Finch, Clara N.
: TITLE OF INVENTION: MUTATIONS RENDERING PLATELET
: TITLE OF INVENTION: GLYCOPROTEIN Ib ALPHA LESS REACTIVE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle
: STREET: Clinton Square, P.O. Box 1051
: CITY: Rochester
: STATE: New York
: COUNTRY: USA
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/119,262B
: FILING DATE: 09-SEP-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/821,717
: FILING DATE: 15-JAN-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Timain, Susan J.
: REGISTRATION NUMBER: 34,103
: REFERENCE/DOCKET NUMBER: 20884/22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716) 263-1636
: TELEFAX: (716) 263-1600
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 610 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PUBLICATION INFORMATION:
: AUTHORS: Lopez, Jose A.
: AUTHORS: Chung, Dominic W.
: AUTHORS: Fujikawa, Kazuo
: AUTHORS: Hagen, Frederick S.
: AUTHORS: Papayannopoulou, Thalia
: AUTHORS: Roth, Gerald J.
: TITLE: Cloning of the alpha chain of human platelet
: TITLE: glycoprotein Ib: A transmembrane protein
: TITLE: leucine-rich alpha-2-glycoprotein
: JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
: VOLUME: 84
: PAGES: 5615-5619

```

TITLE: glycoprotein Ib: A transmembrane protein
 TITLE: leucine-rich alpha-2-glycoprotein
 JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
 VOLUME: 84
 PAGES: 5615-5619

```
Query Match      100.0%; Score 442; DB 3; Length 156;  
Best Local Similarity   100.0%; Pred. No. 1.le-42;  
Matches    86; Conservative     0; Mismatches     0; Indels     0; Gaps     0  
  
Qy       1 VRRQLPVEEPNPKRLLLELLTIIVFCILMAEEGVAPLPPEDADPNAA SLA TPVSPYLE 60  
          | |||||  
Db        71 VRRQLPVEE PN PK R LL E L LT I IV F C IL MA EE GV A P L PP ED AD P N AA S LA TPV SP Y LE 130  
          | |||||  
  
Qy         61 PFNLTSSEUSDYALDLSTFLQQHPAAF 86  
          | |||||  
Db        131 PFNLTSSEPSDYALDLSTFLQQHPAAF 156
```

RESULT 4
US-09-356-952-3
Sequence 3, Application US/09356952
Patent No. 6117663
GENERAL INFORMATION:
APPLICANT: Boriack-Sjodin, Ann
APPLICANT: Margarit, S. M.
APPLICANT: Bor-Sogli, Dafna
APPLICANT: Cole, Phillip
APPLICANT: Kuriyan, John
TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
TITLE OF INVENTION: THEREOF
FILE REFERENCE: 600-1-228N
CURRENT APPLICATION NUMBER: US/09/356,952
CURRENT FILING DATE: 1999-07-19
EARLIER APPLICATION NUMBER: 60/093,631
EARLIER FILING DATE: 1998-07-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1596
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-356-952-3

```

Query Match      18.0%: Score 79.5; DB 3; Length 1596;
Best Local Similarity 28.3%: Pred. No. 1.1;
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps
6;

QY 1 VRRQLPVE-----EPNPKRIILFLLLTVFCQILMAEEGVAPLPDPEDAPNAAASLAPT 53
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1423 TRNSAIKRAAATSOPOAAAGPTSTTLVTVOAVAYDEPLPLPISP-----AKSSSTT 1477

```


Search completed: January 30, 2001, 23:09:04
Job time: 6029 sec

Db 5 KLPLSPSVIRIRILFYLLMLLFCOOLAMIFWRVGLP-----DNSPVASVOITPAQARQQ 58

QY 61 PFNLTSPESDYAL 73
| | | | | : | : |
Matches 29; Conservative 13; Mismatches 39; Indels 21; Gaps 5;

Db 59 PVTL-----NDFTL 67

RESULT 12
S27920
nuclear antigen EBNA-3A - human herpesvirus 4
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000
C:Accession: S27920; G49253; H49253; F49253
R:Sample, J.; Young, L.; Martin, L.; Chatman, T.; Kieff, E.; Rickinson, A.; Kieff, E.
submitted to the EMBL Data Library, July 1990
A:Reference number: S27920
A:Accession: S27920
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-925 <SAM>
A:Cross-references: EMBL:M34440; NID:g330407; PIDN:AAA45893.1; PID:g330408
R:Apolloni, A.; Moss, D.; Stumm, R.; Burrows, S.; Suhrbier, A.; Misko, I.; Schmidt, C.;
Eur. J. Immunol. 22, 183-189, 1992
A:Title: Sequence variation of cytotoxic T cell epitopes in different isolates of Epstein-Barr virus
A:Reference number: A49034; MUID:92111623
A:Accession: G49253
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 314-332 <CAPO>
A:Cross-references: GB:S79235; NID:g242901; PIDN:AAB20990.1; PID:g242902
A:Experimental source: human B-type strain QIMR-JSM6
A:Note: sequence extracted from NCBI backbone (NCBIN:79235, NCBI:P:79252)
A:Accession: H49253
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 314-338 <AP2>
A:Cross-references: GB:S79234; NID:g242899; PIDN:AAB20989.1; PID:g242900
A:Experimental source: human B-type strain AG876JS; human B-type strain L4MR
A:Note: sequence extracted from NCBI backbone (NCBIN:79237, NCBI:P:79254, NCBIN:79234, NCBI:P:79253)
A:Genetics:
A:Introns: 112/3
C:Superfamily: human herpesvirus 4 nuclear antigen EBNA-3A

Query Match 16.6%; Score 73.5; DB 2; Length 925;
Best Local Similarity 28.4%; Pred. No. 17;
Matches 29; Conservative 13; Mismatches 39; Indels 21; Gaps 5;

QY 1 VRRQLPVEPNPAKRLFLLLTIVFCQI--LMAEEGVAPLP-----PEDA 44
| | | | | : | : | | | | |
Db 606 VSPQPMERPLEPCEQMF--PGSPFSQVADVARESGVPAMQPOYFDLPLTQPIISOGAPAA 663

QY 45 PNAASLAPTPVSPVLEP--ENL-TSEPSDYALDLSTFLOQHP 83
| | | | | : | : | | | | |
Db 564 PLRASMGPVPPVPTQPOYFDIPLTEPINOGASAAHFLPOOP 705

RESULT 13
T30351
mucin-like protein - Lymantria dispar nuclear polyhedrosis virus
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T30351
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohrer, J.L.;
Virology 253, 17-34, 1999
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar
A:Reference number: 220836; MUID:99124785
A:Accession: T30351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1029 <KUZ>
A:Cross-references: EMBL:AF081810; PIDN:AAC70189.1

Query Match 16.5%; Score 73; DB 2; Length 1029;
Best Local Similarity 38.2%; Pred. No. 22;
Matches 29; Conservative 3; Mismatches 30; Indels 14; Gaps 4;

QY 6 PVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLP---PEDAPNAASL-----APTVPVSP 57
| | | | | : | : | | | | |
Db 644 PVPEPS-APVDYFTLSAEF-----APEPAPEPAPEAPEPTSRFTSEPALGPVEP 697

QY 58 VLPEFNLTSEPSDYAL 73
| | | | | : | : |
Db 698 ALEPVEPALEPVEPAL 713

RESULT 14
S28013
outC protein - Erwinia chrysanthemi
C:Species: Erwinia chrysanthemi
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C:Accession: S28013; S23885
R:Condemine, G.; Dorel, C.; Hugouvieux-Cotte-Pattat, N.; Robert-Houdouy, J.;
Mol. Microbiol. 6, 3199-3211, 1992
A:Title: Some of the out genes involved in the secretion of pectate lyases in Erwinia chrysanthemi
A:Reference number: S28011; MUID:93086427
A:Accession: S28013
A:Molecule type: DNA
A:Residues: 1-272 <CON>
A:Cross-references: EMBL:X65265; NID:g3152953; PIDN:CAA46369.1; PID:g42201
C:Genetics:
A:Gene: outC
C:Keywords: transmembrane protein

Query Match 16.4%; Score 72.5; DB 2; Length 272;
Best Local Similarity 30.1%; Pred. No. 5.6;
Matches 22; Conservative 16; Mismatches 22; Indels 13; Gaps 4;

QY 4 QLPVEEPNPAKRLFLLLTIVFCQ---ILMAEEGVAPLPPEPADAPNAASLAPTPVSPVLE 60
| | | | | : | : | | | | |
Db 5 KLPLSPSVIRIRILFYLLMLLFCOOLAMIFWRIGL-----PDNAP-VSSVOITPAQARQQ 58

QY 61 PFNLTSPESDYAL 73
| | | | | : | : |
Db 59 PVTL-----NDFTL 67

RESULT 15
NBHUIA
platelet glycoprotein Ib alpha chain precursor - human
N:Alternate names: membrane glycoprotein Ib alpha chain
N:Contains: glycosylated
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 22-Jun-1999
C:Accession: A94174; A60435; A94173; S16945; I55355; A27075; A27102
R:Lopez, J.A.; Chung, D.W.; Fujikawa, K.; Hagen, F.S.; Papayannopoulos, T.; Roth,
Proc. Natl. Acad. Sci. U.S.A. 84, 5615-5619, 1987
A:Title: Cloning of the alpha-chain of human platelet glycoprotein Ib: a transmembrane protein
A:Reference number: A94174; MUID:87289655
A:Accession: A94174
A:Molecule type: mRNA
A:Residues: 1-626 <LOP>
A:Cross-references: GB:J02940; NID:g183499; PIDN:AAA52595.1; PID:g306793
R:Wicki, A.N.; Walz, A.; Gerber-Huber, S.N.; Wenger, R.H.; Vornhagen, R.; Clemet,
Thromb. Haemost. 61, 448-453, 1989
A:Title: Isolation and characterization of human blood platelet mRNA and construction of a cDNA library
A:Reference number: A60435; MUID:90020160
A:Accession: A60435
A:Molecule type: mRNA
A:Residues: 207-467 <WIC>
R:Titani, K.; Takio, K.; Handa, M.; Ruggeri, Z.M.
Proc. Natl. Acad. Sci. U.S.A. 84, 5610-5614, 1987
A:Title: Amino acid sequence of the von Willebrand factor-binding domain of platelet glycoprotein Ib

A:Reference number: Z14130
A:Accession: J00247
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-955 <NAT>
A:Cross-references: EMBL:AB012266; NID:d1227741; PID:BAA32791.1; P1D:d1033575
A:Experimental source: brain
C:Genetics:
A:Gene: wiz

Query Match 16.7%; Score 74; DB 2; Length
Best Local Similarity 41.93; Pred. No. 16;
Matches · 18; Conservative 2; Mismatches 15; In

QY	34	GVPAPLPEDAPNAASLAPTVPSPVLE-----PFNLTSEP	68
		:	
Db	300	GSPTPKNPEDKSPQLSLSPRPTSPRAQCPOSEDEGPLNLISGP	342

RESULT 10
T00246
zinc finger protein wizL - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C:Accession: T00248
R:Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Tohyama, M.
submitted to the EMBL Data Library, March 1998
A:Description: Molecular cloning and distinct developmental expression pattern
A:Reference number: Z14130

A:Accession: J00248
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1561 <MAT>
A:Cross-references: EMBL:AB012265; MID:d1227740; PIDN:BAA32790.1; PID:d1033756
A:Experimental source: brain
C:Genetics:
A:Gene: wiz

Query Match 16.7%; Score 74; DB 2; Length 1561;
Best Local Similarity 41.9%; pred. No. 27;
Matches 18; Conservative 2; Mismatches 15; Indels 8; Gaps 1

Qy	34	GVPAPLPEDAPNAASLAPTVPSPVLE-----PFLNTSEP	68
		: :	
Db	906	GSPTPKNPEDKSPQLSLSPRPTSPRAQCQSEDEGPNLTSGP	948

```

RESULT 11
A47021
    pectic enzyme secretion protein OutC - Erwinia chrysanthemi
C:Species: Erwinia chrysanthemi
C:Date: 24-Feb-1994 #sequence_revision 18-Nov-1994 #text_change
C:Accession: A47021
J:Lindeberg, M.; Collmer, A.
J: Bacteriol. 174, 7385-7397, 1992
A:Title: Analysis of eight out genes in a cluster required for
A:Reference number: A47021; M0ID:93054353
A:Accession: A47021
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-272 <LIN>
A:Note: sequence extracted from NCBI backbone (NCBIP:118270)
C:Keywords: transmembrane protein

```

```

Query Match      16.6%; Score 73.5; DB 2; length 272;
Best Local Similarity 30.1%; Pred. No. 4.5;
Matches 22; Conservative 13; Mismatches 25; Indels 1
y      4 QLPVEEPNPAKRLLELLITIVFC---ILMAEEGVPAPIPPEDPAANASLAPTIV
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

```

QY 61 PF-NLTSEP 68
| | :| |
DB 4657 PLVELPTEP 4665

RESULT 7
AB1141
acyl CoA thioester hydrolase family protein NMB0925 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: AB1141
R:Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Rhee, J.; Jin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.; Science 287, 1809-1815, 2000
A:Authors: Grandl, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: AB1141
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1148 <TET>
A:Cross-references: GR:AE002444; GB:AE002098; NID:g7226162; PIDN:AAF41333.1; PID:g722616
A:Experimental source: serogroup B, strain MMD58
C:Genetics:
A:Gene: NMB0925

```

Query Match      17.2%; Score 76; DB 2; Length 148;
Best Local Similarity 40.5%; Pred. No. 1.3;
Matches 15; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

Qy 18 FLILTIIVFCILMAEEGVAPLPPEADAPNAASLAPTP 54
      ||| : ||| : ||| : ||| : ||| : |||
Db 112 VLTEAVETIVVAIDAEGNRPPIPKENPKLAGLLPTP 148

```

RESULT 8

Tl83344

P-glycoprotein E - Leishmania tropica

C:Species: Leishmania tropica

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000

C:Accession: Tl83344

R:Leifuernte, E.; Castanys, S.; Gamarro, F.
submitted to the EMBL data Library, April 1996

A:Reference number: Z18880

A:Accession: Tl8344

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1677 <LAF>

A:Cross-references: EMBL:U55381; NID:g91916505; PID:g91916506; PIDN:AAB51191.1

C:Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homology

```

Query Match      17.2%; Score 76; DR 2; Length 1677;
Best Local Similarity 33.9%; Pred. No. 19;
Matches 20; Conservative 6; Mismatches 15; Indels 18; Gaps 2;

Qy 37  APLP-----PEDAPNAAASLAPTPVSPVLFPFN-----LTSEPSDYALDI 77
      |||||  : : : : :  |||  : : : : :  |||  : : : : :  |||  : : : : :
Db 698  APLPEAEAFULPDROPSSSAAPRPGAVPTPLNAKSGHTGCGDAAASSEPLSSAAQST 756

```

RESULT 9
T00247
zinc finger protein wiz - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C:Accession: T00247
R:Matsumoto, K.; Ishii, N.; Yoshida, S.; Shimosaka, S.; Wanaka, A.; Tohyama, M.
submitted to the EMBL Data Library, March 1998
A:Description: Molecular cloning and distinct developmental expression pattern of spliced

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:09:03 ; Search time 21.17 Seconds
(without alignments)
275.836 Million cell updates/sec

Title: US-08-799-910-10_COPY_71_156
Perfect score: 442
Sequence: 1 VRRQLPVEEPNPAKRLFL.....EPFDYALDLSTFLOQHPAAF 86

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: pir1.*
- 2: pir2.*
- 3: pir3.*
- 4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	442	100.0	156	2 JC5537	differentiation-de
2	273	61.8	153	2 S33363	gly96 protein - mo
3	81.5	18.4	102	2 T33565	hypothetical prote
4	79.5	18.0	1596	2 A41216	guanine nucleotide
5	78	17.6	4957	2 T03455	ALR protein - huma
6	78	17.6	5262	2 T03454	ALR protein - huma
7	76	17.2	148	2 A81141	acyl CoA thioester
8	76	17.2	1677	2 T18344	P-glycoprotein E -
9	74	16.7	955	2 T00247	zinc finger protei
10	74	16.7	1561	2 T00248	zinc finger protei
11	73.5	16.6	272	2 A47021	pectic enzyme secr
12	73.5	16.6	925	2 S27920	nuclear antigen EB
13	73	16.5	1029	2 T30351	mucin-like protein
14	72.5	16.4	272	2 S28013	outC protein - Erw
15	71.5	16.2	626	1 NBHUA	platelet glycoprot
16	71	16.1	359	2 T33756	hypothetical prote
17	70.5	16.0	235	2 D83477	hypothetical prote
18	70.5	16.0	605	1 Q0BE29	BRLFI protein - hu
19	70	15.8	148	2 E81878	probable acyl-CoA
20	70	15.8	1233	2 T37045	nitrate reductase
21	69.5	15.7	1887	2 S61703	fatty-acid synthas
22	69	15.6	418	2 T19800	hypothetical prote
23	69	15.6	1402	2 I46707	translation initia
24	69	15.6	2187	2 T30826	nascent polypeptid
25	68.5	15.5	181	2 T08793	hypothetical prote
26	68.5	15.5	882	1 IJHUCF	cadherin 1 precurs
27	68.5	15.5	1072	2 T50949	verprolin related
28	68.5	15.5	1337	1 I38670	protein-tyrosine-p
29	68.5	15.5	1844	2 S01956	hypothetical prote

30	68.5	15.5	2649	2 T51023	hypothetical prote
31	68	15.4	241	2 T22216	hypothetical prote
32	68	15.4	267	2 A49068	cranosynotosis-as
33	68	15.4	1396	2 A44453	translation initia
34	67	15.2	283	2 S13383	hydroxyproline-ric
35	67	15.2	586	2 S19381	hypothetical prote
36	67	15.2	3566	1 A40701	tenascin-x precurs
37	66.5	15.0	169	2 A72466	hypothetical prote
38	66.5	15.0	3149	1 Q0BE8	BRLFI protein - hu
39	66	14.9	101	2 T46506	hypothetical prote
40	66	14.9	187	2 C75558	acetyl-CoA carboxy
41	66	14.9	621	2 JC5164	acetylacate synth
42	66	14.9	5149	2 F83345	probable non-ribos
43	65.5	14.8	325	2 D70666	exBB protein - pse
44	65.5	14.8	329	2 S28442	hypothetical prote
45	65.5	14.8	424	2 T43468	

ALIGNMENTS

RESULT 1

JC5537
differentiation-dependent protein DIF-2 - human
C:Species: Homo sapiens (man)
C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 07-May-1999
C:Accession: JC5537
R:Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.
Biochem. Biophys. Res. Commun. 235, 4-9, 1997
A:Title: Identification and characterization of a novel monocyte/macrophage diff
A:Reference number: JC5537; MUID:97339426
A:Accession: JC5537
A:Molecule type: mRNA
A:Residues: 1-156 <PIE>
A:Experimental source: monocyte
A>Note: the authors translated the codon CCG for residue 106 as Arg

Query Match 100.0%; Score 442; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.2e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VRRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPPEADPNAAASLAPTVPSPVLE 60
Db	71	VRRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPPEADPNAAASLAPTVPSPVLE 130
Qy	61	PFNLTSEPSDYALDLSTFLOQHPAAF 86
Db	131	PFNLTSEPSDYALDLSTFLOQHPAAF 156

RESULT 2

S33363
gly96 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C:Accession: S33363
R:Charles, C.H.; Yoon, J.K.; Simske, J.S.; Lau, L.F.
Oncogene 8, 797-801, 1993
A:Title: Genomic structure, cDNA sequence, and expression of gly96, a growth fac
A:Reference number: S33363; MUID:93173526
A:Accession: S33363
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-153 <CHA>
A:Cross-references: EMBL:X67644
C:Genetics: 70/3
C:Introns: 70/3
C:Keywords: transmembrane protein

Query Match 61.8%; Score 273; DB 2; Length 153;
Best Local Similarity 68.2%; Pred. No. 1.3e-19;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 21:28:30 ; Search time 49.1 seconds
(without alignments)
846.830 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468
Perfect score: 258
Sequence: 1 GTCCGGCCCGACGTCAGT.....AGCAACACCCGGCCGCTTC 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues
Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/PCUS_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	100.0	1228	3	US-08-826-246-9
2	258	100.0	1228	3	US-08-944-495-9
3	258	100.0	1228	3	US-09-126-640-5
4	43	16.7	7218	1	US-08-232-463-14
5	39.4	15.3	2847	1	US-08-087-007-2
6	39.4	15.3	2847	3	US-08-483-433-2
7	39.4	15.3	2847	4	PCT-US92-05920-2
8	38.6	15.0	6354	3	US-09-058-389A-5
9	37.6	14.6	170	3	US-09-058-389A-13
10	34.8	13.5	2953	2	US-08-859-201-1
11	33.6	13.0	1120	1	US-08-188-582-8
12	33.6	13.0	1120	1	US-08-846-715-8
13	32	12.4	3765	3	US-07-705-490-1
14	32	12.4	4362	2	US-08-455-073A-1
15	31.8	12.3	599	1	US-08-584-226-3
16	31.6	12.2	2074	2	US-09-018-576-2
17	31.6	12.2	2074	2	US-09-018-576-11
18	31.6	12.2	2074	3	US-09-248-137-2
19	31.6	12.2	2074	3	US-09-248-137-11
20	31.4	12.2	1335	4	PCT-US91-06532-1
21	31.4	12.2	38506	3	US-09-320-878-19
22	30.8	11.9	797	1	US-08-332-467B-2
23	30.8	11.9	797	1	US-08-681-811-2
24	30.8	11.9	797	4	PCT-US93-12507-2
25	30.8	11.9	2680	3	US-09-063-035-1
26	30.8	11.9	7791	3	US-08-949-386-23
27	30.6	11.9	7791	3	US-08-450-562-23
28	30.6	11.9	7808	3	US-08-949-386-22

C 29 30.6 11.9 7808 3 US-08-450-562-22
30 30.2 11.7 22481 4 PCT-US95-07201-43
31 30 3688 5 5248670-4
32 29.8 11.6 536 1 US-08-329-704-1
33 29.8 11.6 536 2 US-08-472-604-1
34 29.8 11.6 536 2 US-08-486-117-1
35 29.8 11.6 538 2 US-08-865-337A-3
36 29.8 11.6 2196 2 US-08-865-337A-2
37 29.8 11.6 8299 1 US-08-462-014-2
38 29.8 11.6 8299 3 US-08-923-137-3
39 29.6 11.5 693 1 US-08-168-091A-3
40 29.6 11.5 1219 4 PCT-US93-06251-11
41 29.6 11.5 1313 1 US-08-176-427B-7
42 29.6 11.5 1313 2 US-08-356-060A-4
43 29.6 11.5 1313 3 US-08-460-900C-4
44 29.6 11.5 1489 3 US-08-836-582-1
45 29.6 11.5 2485 1 US-08-424-424B-1

ALIGNMENTS

RESULT 1
US-08-826-246-9
Sequence 9, Application US/08826246
Patent No. 6048709
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE TREATMENT AND DIAGNOSIS OF
CARDIOVASCULAR DISEASE
TITLE OF INVENTION: 44
NUMBER OF SEQUENCES:
CORRESPONDENCE ADDRESS:
ADDRESSEE: PERNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,246
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011,787
FILING DATE: 16-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-078-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1228 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...468
OTHER INFORMATION:

QY 241 CAACACCGCGCCGCTTC 258
|||||
Db 346 caacacccggccgccttc 363

RESULT 7

```

US:09-522-303-1025
: Sequence 1025, Application US:09522303
: GENERAL INFORMATION:
: APPLICANT: Gearing, David P.
: APPLICANT: Holtzman, Douglas A.
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
: TITLE OF INVENTION: HUMAN MAMMARY EPITHELIUM LIBRARY
: FILE REFERENCE: 1600.1086-001
: CURRENT APPLICATION NUMBER: US:09/522.303
: CURRENT FILING DATE: 2000-03-08
: EARLIER APPLICATION NUMBER: 60/123.393
: EARLIER FILING DATE: 1999-03-08
: NUMBER OF SEQ ID NOS: 1353
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1025
: LENGTH: 657
: TYPE: DNA
: ORGANISM: Homo sapiens
US:09-522-303-1025

```

Query Match	100.0%;	Score 258;	DB 19;	Length 657;
Best local Similarity	100.0%;	Prod. No. 7.6e-52;		
Matches 258;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps				
QY	1	GTCCGGCGCCAGCTGCCAGTCGAGGAAACCAGCACCCAGCAAAAGGCTTCCTTTCTGCTG	60	
Db	277	gtcggcgccagctgcccagtcggaggaaaccgagccagccaaagagcttcttctctgctg	336	
QY	61	CTCACCATCGTCTTCGCCAGATCTCGTATGCTCGTAGAGGGTGTCGCGGGCGCCCTGCCCT	120	
Db	337	ctcaccatcgctctctgcagatcctgtagtgcgtgaaggagggtgtagcgcgccctgcct	396	
QY	121	CCAGAGAGACGCCCTCTAACGCCCATCTCCTTGGCGCCACCCCTGTGTCGCCCGTCTCGAG	180	
Db	397	ccagagagcgcgccttaacgcgcacccctgcggccacccctgtgtccccgcctctcgag	456	
QY	181	CCCTTTTAATCTGACTTTCGGAGCCCTCGGACCTACGCTCTGGACCTCAGACCTTTCCTCCAG	240	
Db	457	cccltctaactctgacttcggagccctcgactacgctggacctcagcacttctctccag	516	
QY	241	CAACACCCGCGCCCTTC	258	
Db	517	caacacccggcgcttc	534	

RESULT

US-09-698-010-14935
Sequence 14935, Application US/09698010
GENERAL INFORMATION:
APPLICANT: Williamson, Mark
APPLICANT: Shvjan, Andrew W.
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
TITLE OF INVENTION: THEREFOR
FILE REFERENCE: 1600.2029-001
CURRENT APPLICATION NUMBER: US/09/698,010
CURRENT FILING DATE: 2000-10-27
PRIOR APPLICATION NUMBER: 60/162,358
PRIOR FILING DATE: 1999-10-29
NUMBER OF SEQ ID NOS: 15684
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14935
LENGTH: 673
TYPE: DNA
ORGANISM: Homo sapiens
IS-09-698-010-14935

Query Match	100.0%	Score	258;	DB	23;	Length	673;
Best Local Similarity	100.0%;	Pred. No.	7.7e-02;				
Matches	258;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						

Oy	1	GNCCGGCGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGAGCTTTCTCTTTCTGCTG	60
Db	269	glccggcgccagctgcccagtcgaggaacgcgaacccacgacaaaggctctcttctctgctg	328
Oy	61	CTCACCATCTGCTTCTGTGCGACAGATCGTGTGCTGAAGAGGGTGTGCCGGCGCCCTTGCT	120
ub	329	ctcaccatcgctctcttcgcagatcctgatgctgaaggagggtgtgcgcygcacctgacct	388
Oy	121	CCAGAGGACGGCCCTTAACGCGCGCATCCCTGGCGCCACCCCTGTCTGCCCGTCTCTCGAG	180
Db	389	ccagagagcgcccttaacgcgcgcatccctcgcgccaccctctgctcccccgctctcgag	448
Oy	181	CCCTTTTAATCTGACTCTGGAGCCCTTCGGGACTACGCTCTTGGACCTCAGCACATTTCTCCAG	240
Db	449	ccctttaatctgactctggagccctcgagctaactctggaaccttcgaccttctctccag	508
Oy	241	CAACACCGCGCGCGCTTC	258
Db	509	caacacccgcgcgcttc	526

RESULT

```

US-09-699-998-9797
; Sequence 9797, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 9797
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-699-998-9797

```

Query Match	100.0%;	Score 258;	DB 23;	Length 673;
Best Local Similarity	100.0%;	Pred. No. 7.7e-52;		
Matches 258;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps				
QY	1	GTCCGGGGCAGTCGCAGTCGAGGAACCGAACCCAGCCAAAGCGCTTCTCTTTCTGCTG	60	
Db	269	gtccggcgccagctgcagctcgaggaaacggaacccagcgaaggtctcttctctg	328	
QY	61	CTCACCATCGTCTCTGCCAGATCTCATGGCTAAAGAGGdTGTCGGCGGCCCGCTGCCT	120	
Db	329	ctaccatcgtctctcgccagatcctctgctgaggagggtgfcgcygcyccctgcct	388	
QY	121	CCAGAGAGCGCCCTAACGCCGATCCCTCGCGCCACCCCTGTGTCCCGGTCCTCGAG	180	
Db	389	ccagaggacgcccctaacgcgcgcctccctcggccaccacctgctcccccgctctcag	448	
QY	181	CCCTTTTAATCTGACTTCGGAGGCCCTCGGACTACGGCTGTGGACCTTCAGCACCTTCTCCACG	240	
Db	449	ccctttaaactgactctcgagccctcggactacgctctggaacctcagacttctctccag	508	
QY	241	CAACACCCGGCGCGCTTC	258	
Db	509	caacacccggcgcccttc	526	


```

RESULT 10
US-09-710-286-3772
; Sequence 3772. Application US/09710286
; GENERAL INFORMATION:
; APPLICANT: Gearling, David P.
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2005-001
; CURRENT APPLICATION NUMBER: US/09/710,285
; CURRENT FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: 60/164,255
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4115
; SOFTWARE: FastSEQ for Windows version 4.0
; SEQ ID NO 3772
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-710-286-3772

```

	Query Match	100.0%	Score 258:	DB 24:	Length 673:
	Best Local Similarity	100.0%	Pred. No. 7,7e-52:		
	Matches 258:	Conservative	0:	Mismatches 0:	Indels 0:
	Gaps				
QY	1	GTCCGGCCCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCGTG	60		
DB	269	gtccggcgccagctgccagtcgaggaaccgaaaccgacaaagcctctctctctctgctg	328		
QY	61	CTACCATCTGCTTCTTGCCAGATCTGATGGCTGAAGGGTGTCGCGCGCCGCTTGGCCT	120		
nb	329	ctcaccatctgtcttcgcagatcctga tggctggaagggtgctgccggcgccctgacct	388		
QY	121	CCAGAGGAGCGCCCTAACGCCGATCCCTGGCGCCACCCCTGTGTCCCCGCTCTCGAG	180		
DB	389	ccagaggacgcccctcaagcgcgcatccctggcgccaccacctgtgtcccccgctctcgag	448		
QY	181	CCCTTTTAATCTGACTTCGGAGGCCCTCGGACTACCGCTCTGGACCTTTCCTCCAG	240		
DB	449	ccctllaatctgacttcggagccctcgactaagcttgagacctcagcattctcccaag	508		
QY	241	CAACACCCGCCGCCCTTTC	258		
DB	509	caaccacggcgcccttc	526		

```

RESULT 11
US-09-716-473-2500
: Sequence 2500, Application US/09716473
: GENERAL INFORMATION:
: APPLICANT: Williamson, Mark
: APPLICANT: Shyjan, Andrew W.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600.2043-001
: CURRENT APPLICATION NUMBER: US/09/716,473
: CURRENT FILING DATE: 2000-11-20
: PRIOR APPLICATION NUMBER: 60/166,502
: PRIOR FILING DATE: 1999-11-19
: NUMBER OF SEQ ID NOS: 2933
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2500
: LENGTH: 673
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-716-473-2500

```

Query Match	100.0%;	Score 258;	DB 24;	Length 673;
Best local Similarity	100.0%;	Pred. No. 7.7e-52;		

```

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 GTCCGGCGCCAGCTCCAGTCGAGGAACCGAACCCAGCCAAAGCCCTTCCTTCCTGC 60
    |||||
DB 269 gtccggcgccagctgcagtcgaggaaccgaaaccagccaaaggtctcttctctg 328
    |||||
QY 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGGCCCTGCCT 120
    |||||
DB 329 ctcaccatctctctgccagatcctggtctgagaggtgtgcggcgccctgct 388
    |||||
QY 121 CCAGGAGAGCGCCCTAACCGCGCATCCCTGCGGCCACCCCTGCTGCCCTCCCTCGAG 180
    |||||
DB 389 ccagaggagcccttaacgcgcctccctggcgccacccctgtgcccccgacctc 448
    |||||
QY 181 CCCTTTAACTCTGACTTCGGAGCGCTCGGACTACGCTCTGGAGCTCAGCACCTTCCTCCAG 240
    |||||
DB 449 ccctttaactctgacttcggagccctcgagactcgtctggacctcagcacttctccag 508
    |||||
QY 241 CAACACCCCGCGCCCTC 258
    |||||
DB 509 caacacccgcgccctc 526
    |||||

RESULT 12
US-09-721-588-4548
; Sequence 4548, Application US/09721588
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Villeval, Jean-Luc
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2046-001
; CURRENT APPLICATION NUMBER: US/09/721,588
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: 60/167,381
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 5410
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 4548
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-721-588-4548

```

```

Query Match      100.0%; Score 258; DB 24; length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

yy      1  GTCCGGCGCCAGCTGCCAGTCGAGGAACCCAGCCAAAGGCTTCGCTTCTCTGCTG 60
          |||
bb      269  gtccggcgccagctgcagtcgaggacccgaccccgccaaaggctctcttcttctgctg 328
          |||

yy      61  CTCACCATCTCTCTGCCAGATCCTGATGGTGTGAAGAGGGTGTGCCGGCGCCCTGCCT 120
          |||
bb      329  ctcaccatctcttcttccagatccctgatactgataagagggtgtgcggcgccctgcct 388
          |||

yy      121  CCAGAGAGCGCCCTAAACGGCCGATCCCTGGCGCCACACCCCTGTGTCGCCCGTCCTCGAG 180
          |||
bb      389  ccagagagcgcctctaagccgcatccctggcgcccaacccclgtgtcccccgttcctcgg 448
          |||

yy      181  CCCTTTAACTCTGACTTCGGAGCCCTCGGACTTACGCTCTGGAGCCTCAGACACTTTCCTCCAG 240
          |||
bb      449  ccccttaactgacttcggagccctcggaactagctctggacctcgacacttctccctcag 508
          |||

yy      241  CAACACCGCGCGCGCTTC 258
          |||
bb      509  caacacccgacgcgcttc 526
          |||

```

RESULT 13
US-09-726-787-2982

; Sequence 2982, Application US/09726787
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2010-001
; CURRENT APPLICATION NUMBER: US/09/726,787
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/168,132
; PRIOR FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 3241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2982
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-787-2982

Query Match 100.0%; Score 258; DB 55; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCCGAGTGCAGTCGAGGAAACCGAACCCCAAGAGGCTTCTTTCTGCTG 60
Db 269 gtccggccgagctgcagtcgaggaacccgaaacccgaaagcttcttcttctg 328

QY 61 CTCACCATCTCTTCTCCAGATCCCTGATGGCTGAAGAGGGTGTGCGGGCCCTGCCT 120
Db 329 ctccaccatctcttctccagatccctgatggctgaagagggtgtgccccctgcct 388

QY 121 CCAGAGACCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGCTCCCGCTCCGAG 180
Db 389 ccagagagcccttaacgccgatccctggcgccacccctgtgccccctgcctgag 448

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTTCTCCACG 240
Db 449 ccctttaatctgacttcggagccctcggactacgctctggaacctcagcacttctccacg 508

QY 241 CAACACCCGGCGCCTTC 258
Db 509 caacacccggcgcccttc 526

RESULT 14
US-09-699-998-1778
; Sequence 1778, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1778
; LENGTH: 706
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(706)
; OTHER INFORMATION: n = A,T,C or G
US-09-699-998-1778

Query Match 100.0%; Score 258; DB 23; Length 706;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCCGAGTGCAGTCGAGGAAACCGAACCCCAAGAGGCTTCTTTCTGCTG 60
Db 195 gtccggccgagctgcagtcgaggaacccgaaacccgaaagcttcttcttctgclg 254

QY 61 CTCACCATCTCTTCTCCAGATCCCTGATGGCTGAAGAGGGTGTGCGGGCCCTGCCT 120
Db 255 ctccaccatctcttctccagatccctgatggctgaagagggtgtgccccctgcct 314

QY 121 CCAGAGACCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGCTCCCGCTCCGAG 180
Db 315 ccagagagcccttaacgccgatccctggcgccacccctgtgccccctgcctcag 374

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTTCTCCACG 240
Db 375 ccctttaatctgacttcggagccctcggactacgctctggaacctcagcacttctccacg 434

QY 241 CAACACCCGGCGCCTTC 258
Db 435 caacacccggcgcccttc 452

RESULT 15
US-09-699-998-10469
; Sequence 10469, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10469
; LENGTH: 803
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-699-998-10469

Query Match 100.0%; Score 258; DB 23; Length 803;
Best Local Similarity 100.0%; Pred. No. 7.8e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCCGAGTGCAGTCGAGGAAACCGAACCCCAAGAGGCTTCTTTCTGCTG 60
Db 184 gtccggccgagctgcagtcgaggaacccgaaacccgaaagcttcttcttctgclg 243

QY 61 CTCACCATCTCTTCTCCAGATCCCTGATGGCTGAAGAGGGTGTGCGGGCCCTGCCT 120
Db 244 ctccaccatctcttctccagatccctgatggctgaagagggtgtgccccctgcct 303

QY 121 CCAGAGACCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGCTCCCGCTCCGAG 180
Db 304 ccagagagcccttaacgccgatccctggcgccacccctgtgccccctgcctcag 363

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTTCTCCACG 240
Db 364 ccctttaatctgacttcggagccctcggactacgctctggaacctcagcacttctccacg 423

QY 241 CAACACCCGGCGCCTTC 258
Db 424 caacacccggcgcccttc 441

Search completed: January 30, 2001, 22:04:50
Job time: 13154 sec

ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949rls
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA USA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455.073A
FILING DATE: 31-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: UPN-2201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4362 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..2118
US-08-455-073A-1

Query Match 12.4%; Score 32; DB 2; Length 4362;

Best Local Similarity 49.5%; Pred. No. 4.2;
Matches 109; Conservative 0; Mismatches 110; Indels 1; Gaps 1;
QY 28 CGCAACCCAGCCAAAGGCTTCTTCTGCTGCTCACCATCGTCTTCTGCGCAGATCTCG 87
DB 246 CGGCATTCACACACACAGCTCTCCATCTTCTTCAGCCCTGCTAGCCCGGAGCCCG 187
QY 88 ATGGCTCAAGAGGGTGTGCGCGCCCTTCCCTCCAGAGAGCCGCCCTTAACCGCCATCC 147
DB 186 CCCCCAGAGGTGGTGGCGGCGCTCGAGGC-CCAGCGCGCGCGCGCGCGCGCGCG 128
QY 148 CTGGCGCCACCCCTGTGTCCTCCCGCTCTCGAGCCCTTTAATCTGACTTCGAGCCCTCG 207
DB 127 GCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCG 68
QY 208 GACTAGCTCTGACCTCAGCACTTCTCTCCAGCAACCC 247
DB 67 GCGCGCGCGCTCG 28

RESULT 15

US-08-584-226-3/c
Sequence 3, Application US/08584226
Patent No. 5798240

GENERAL INFORMATION:

APPLICANT: Martinis, Susan A.
APPLICANT: Sassanfar, Mandana
APPLICANT: Kim, Sunghoon
APPLICANT: Lee, Sang Ho
APPLICANT: Schimmel, Paul R.
TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL METHIONYL-TRNA
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington

STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/584,226
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/305,766
FILING DATE: 13-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI94-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 599 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-584-226-3

Query Match 12.3%; Score 31.8; DB 1; Length 599;

Best Local Similarity 48.1%; Pred. No. 3;

Matches 90; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 65 CCATCGTCTTCTGCCAGATCCTTGATGGCTGAAGAGGGTGTGCGCGCGCCCTTGCTCCAG 124
DB 591 CCGCGGAATTCTATCTGACGGGCGGATGAGCAGCGCTGAAGTGGCGAGGCGCGCG 532
QY 125 AGGAGCGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAGCCCT 184
DB 531 CGCGCGCGCGGCTGCGCGCACAGCTTGGCCCGCGCAATTCCGACGCTTCCACCGCA 472
QY 185 TTAATCTGACTTCGGAGCGCCCTCGGACTACGCTCTGGACTCAGCACTTCTCTCCAGCAAC 244
DB 471 TGCAGGAGCGCTGACATCTCTCTCGACCGGTTTCATCCGACACCCCGGACCACT 412
QY 245 ACCCGGC 251
DB 411 ACAGCGC 405

Search completed: January 30, 2001, 21:28:34
Job time: 19996 sec

RESULT 13
US-07-705-490-1/c
: Sequence 1, Application US/07705490
: Patent No. 6107025
: GENERAL INFORMATION:
: APPLICANT: Caskey, C. T.
: APPLICANT: Neilson, David L.
: APPLICANT: Pieretti, Maura

```

US-08-453-073A-1/c
; Sequence 1, Application US/08455073A
; Patent No. 5876949
;
; GENERAL INFORMATION:
; APPLICANT: Gideon Dreyfuss
; APPLICANT: Mikiko C. Siomi
; APPLICANT: Yan Zhang
; TITLE OF INVENTION: Fragile X Re
; TITLE OF INVENTION: Of Making An
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:

```


Result No.	Query		Length	DB	ID	Description
	Score	Match				
1	442	100.0	156	2	JC5537	differentiation-de
2	273	61.8	153	2	S3363	gly96 protein - mo
3	81.5	18.4	153	2	S3363	hypothetical prote
4	79.5	18.0	1596	2	A41216	guanine nucleotide
5	78	17.6	4957	2	T03455	ALR protein - huma
6	78	17.6	5262	2	T03454	ALR protein - huma
7	76	17.2	148	2	A81141	acyl CoA thioester
8	76	17.2	1677	2	T18344	p-glycoprotein E -
9	74	16.7	955	2	T00247	zinc finger protei
10	74	16.7	1561	2	T00248	zinc finger protei
11	73.5	16.6	272	2	A47021	pectic enzyme secr
12	73.5	16.6	925	2	S27920	nuclear antigen EB
13	73	16.5	1029	2	T30351	mucin-like protein
14	72.5	16.4	272	2	S28013	outC protein - Erw
15	71.5	16.2	626	1	NBH01A	platelet glycoprot
16	71	16.1	359	2	T33756	hypothetical prote
17	70.5	16.0	235	2	D83477	hypothetical prote
18	70.5	16.0	605	1	Q0BE29	BRFL1 protein - hu
19	70	15.8	148	2	E81878	probable acyl-CoA
20	70	15.8	1233	2	T37045	nitrate reductase
21	69.5	15.7	1887	2	S61703	fatty-acid synthas
22	69	15.6	418	2	T19800	hypothetical prote
23	69	15.6	1402	2	I46707	translation initia
24	69	15.6	2187	2	T30826	nascent polypeptid
25	68.5	15.5	181	2	T08793	hypothetical prote
26	68.5	15.5	882	1	IJHUCR	cadherin I precurs
27	68.5	15.5	1072	2	T50949	proteolin related
28	68.5	15.5	1337	1	I138670	protein-tyrosine-p
29	68.5	15.5	1844	2	S01956	hypothetical prote

Query Match	61.8%;	Score 273;	DB 2;	Length 153;
Best Local Similarity	68.2%;	Pred. No. 1.3e-19;		

QY 61 PF-NLTSEP 68
| | | | |
Db 4657 PLVELPTPEP 4665

RESULT 7
A81141
acyl CoA thioester hydrolase family protein NMB0925 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: A81141
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.; Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vevea, A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: A81141
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-148 <TET>
A:Cross-references: GB:AE002444; GB:AE002098; NID:g7226162; PIDN:AAF41333.1; PID:g722616
A:Experimental source: serogroup B, strain MMD58
C:Genetics:
A:Gene: NMB0925

Query Match 17.2%; Score 76; DB 2; Length 148;
Best Local Similarity 40.5%; Pred. No. 1.3;
Matches 15; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 18 FLLTIVFCOILMAEGVPAPLPDPEDAPNAASLAPTP 54
| | | | | | | | | | | | | | | | | | | | | |
Db 112 YLVEAVTVVAIDEGNPRIPKEGNPKLAGLLPTP 148

RESULT 8
T18344
P-glycoprotein E - Leishmania tropica
C:Species: Leishmania tropica
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T18344
R:Lafuente, E.; Castanys, S.; Gamarro, F.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z18880
A:Accession: T18344
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1677 <LAF>
A:Cross-references: EMBL:U55381; NID:g1916605; PID:g1916606; PIDN:AAB51191.1
C:Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homology

Query Match 17.2%; Score 76; DB 2; Length 1677;
Best Local Similarity 33.9%; Pred. No. 19;
Matches 20; Conservative 6; Mismatches 15; Indels 18; Gaps 2;

QY 37 APLP-----PEDAPNAASLAPTPVSPVLEPFP-----LTSEPSDYALDIIST 77
| | | | | | | | | | | | | | | | | | | | | |
Db 698 APLPEAEAFIPDROPPSSSSAAPRPAGVPVTEPLNAGSHGTGDAASSEPLSSAQKST 756

RESULT 9
T00247
zinc finger protein wiz - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C:Accession: T00247
R:Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Tohyama, M.
submitted to the EMBL Data Library, March 1998
A:Description: Molecular cloning and distinct developmental expression pattern of spliced

Search completed: January 30, 2001, 23:09:04
Job time: 6029 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:06:25 ; Search time 36.91 Seconds
(without alignments)
41.840 Million cell updates/sec

Title: US-08-799-910-10_COPY_71_156

Perfect score: 442

Sequence: 1 VRRQLPVEPNPAKRLLELL.....EPSDYALDLSTFLQHPAAAF 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep:*

2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep:*

3: /cgn2_6/ptodata/2/1aa/6-COMB.pep:*

4: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep:*

5: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	442	100.0	156	3	US-08-826-246-10
2	442	100.0	156	3	US-08-944-495-10
3	442	100.0	156	3	US-09-126-640-11
4	79.5	18.0	1396	3	US-09-356-952-3
5	71.5	16.2	610	1	US-07-821-717B-6
6	71.5	16.2	610	1	US-08-119-262B-6
7	71.5	16.2	610	1	US-08-135-929A-11
8	71.5	16.2	610	1	US-08-234-265A-11
9	69.5	15.7	739	3	US-09-035-648-24
10	68.5	15.5	878	1	US-08-237-919-2
11	68.5	15.5	878	1	PCT-US95-05518-2
12	68.5	15.5	1337	3	US-08-854-585-2
13	68.5	15.5	1337	3	PCT-US95-05512-2
14	65.5	14.8	286	1	US-08-382-184-3
15	65.5	14.8	286	1	US-08-641-356-3
16	65.5	14.8	325	1	US-08-382-184-2
17	65.5	14.8	325	1	US-08-641-356-2
18	65.5	14.8	1093	3	US-08-545-860D-55
19	65.5	14.8	1093	4	PCT-US94-04496-55
20	65	14.7	328	2	US-08-651-818A-2
21	64.5	14.6	147	2	US-08-530-165-2
22	64.5	14.6	147	2	US-08-530-165-3
23	64.5	14.6	863	2	US-08-666-271-2
24	63.5	14.4	211	2	US-08-164-292B-2
25	63.5	14.4	211	3	US-08-845-623-2
26	63.5	14.4	211	3	US-08-815-927-2
27	63	14.3	174	2	US-08-683-262B-47
28	62	14.0	652	1	US-08-318-831-8

29 61.5 13.9 166 1 US-07-935-311A-2
30 61.5 13.9 166 1 US-08-368-079-2
31 61.5 13.9 166 4 PCT-US93-07996-2
32 61 13.8 297 2 US-08-580-545B-6
33 61 13.8 297 3 US-09-262-653A-6
34 61 13.8 306 2 US-08-824-707-2
35 61 13.8 486 2 US-08-942-423-3
36 61 13.8 537 3 US-08-920-610-2
37 60.5 13.7 259 3 US-08-469-318-133
38 60.5 13.7 259 3 US-08-468-609A-133
39 60.5 13.7 259 4 PCT-US95-01185-133
40 60.5 13.7 347 2 US-09-004-502-1
41 60.5 13.7 1297 2 US-08-290-731C-4
42 59.5 13.5 259 3 US-08-469-318-131
43 59.5 13.5 259 3 US-08-469-318-132
44 59.5 13.5 259 3 US-08-468-609A-131
45 59.5 13.5 259 3 US-08-468-609A-132

ALIGNMENTS

RESULT 1
US-08-826-246-10
Sequence 10, Application US/08826246
Patent No. 6048709
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE TREATMENT AND DIAGNOSIS OF
TITLE OF INVENTION: CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,246
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011,787
FILING DATE: 16-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-078-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-826-246-10

; FILING DATE: 16-NOV-1988
 ; PUBLICATION DATE: 24-MAY-1989
 ; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293
 US-07-821-717B-6

Query Match	16.2%	Score 71.5;	DB 1;	Length 610;
Best Local Similarity	45.0%	Pred. No. 2.7;		
Matches 18;	Conservative	2;	Mismatches	11;
			Indels	

```

Qy 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFLTSEPS 69
    | | | | | | | | | | | | | | | | | | | |
Db 366 PTSPPTTSEPVPEPAPNNTLTETPTSPPTTEP---TSEPA 402
    | | | | | | | | | | | | | | | | | | | |

```

RESULT 6
US-08-119-262B-6
: Sequence 6, Application US/08119262B

Patent No. 5492809

GENERAL INFORMATION:

APPLICANT: Miller, Jonathan L.

APPLICANT: Cunningham, David

APPLICANT: Lyte, Vicki A.

APPLICANT: Finch, Clara N.

```

: TITLE OF INVENTION: MUTATIONS RENDERING PLATELET
: TITLE OF INVENTION: GLYCOPROTEIN IB ALPHA LESS REACTIVE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
:
:
:

```

ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA

ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,262B
; FILING DATE: 00-SEP-1992

```

DATE: 03 SEP 1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/821,717
FILING DATE: 15-JAN-1992

ATTORNEY/AGENT INFORMATION:
NAME: Timaln, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20884/22

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716) 263-1636
: TELEFAX: (716) 263-1600
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:

```

```

;
; LENGTH: 610 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;

```

MOLECULE TYPE: protein
PUBLICATION INFORMATION:
AUTHORS: Lopez, Jose A.
AUTHORS: Chung, Dominic W.
AUTHORS: Fujikawa, Kazuo

AUTHORS: Fujikawa, Kazuo
 AUTHORS: Hagen, Frederick S.
 AUTHORS: Papayannopoulos, Thalia
 AUTHORS: Roth, Gerald J.
 TITLE: Cloning of the alpha chain of human platelet

TITLE: glycoprotein Ib: A transmembrane protein
 TITLE: leucine-rich alpha-2-glycoprotein
 JOURNAL: PROC. NATL. ACAD. SCI. U.S.A.
 VOLUME: 84
 PAGE: 5615-5619

PAGES: 3013-3014

```

:
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson, P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/035,648
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/818,829
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Meiklejohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 07334/003001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-542-5070
: TELEFAX: 617-542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 739 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: US-09-035-648-24
:
: Query Match 15.7% Score 69.5; DB 3; Length 739;
: Best Local Similarity 37.8%; Pred. No. 5,6;
: Matches 17; Conservative 8; Mismatches 19; Indels 1; Gaps 1:
:
: QY 25 FCQILMAEGVPAPLPEDAPNAASLAP-TPVSPVLEPFLNTSEP 68
: | | | | | | | | | | | | | | | | | | | | | |
: Db 391 FMOSSSAKQTPPPVAPKPAVKSSSQPPTVSPVMSQVAPTQP 435
:
: RESULT 10
: US-08-237-919-2
: Sequence 2, Application US/08237919
: Patent No. 5610281
: GENERAL INFORMATION:
: APPLICANT: Brenner, Michael B
: ATTORNEY/AGENT INFORMATION:
: NAME: Cepek, Karyn L.
: TITLE OF INVENTION: Methods and Compositions for
: Modulating Heterotypic E-cadherin Interactions with T Lymphocytes
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wolf, Greenfield & Sacks P.C.
: STREET: 600 Atlantic Avenue
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/237,919
: FILING DATE:
: CLASSIFICATION: 424
:
: Query Match 15.5% Score 68.5; DB 4; Length 878;
: Best Local Similarity 25.7%; Pred. No. 9;
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Plumer, Elizabeth R.
: REGISTRATION NUMBER: 36,637
: REFERENCE/DOCKET NUMBER: B0801/7023
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-720-3500
: TELEFAX: 617-720-2441
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 878 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-237-919-2
:
: Query Match 15.5% Score 68.5; DB 1; Length 878;
: Best Local Similarity 25.7%; Pred. No. 9;
: Matches 26; Conservative 11; Mismatches 31; Indels 33; Gaps 5;
:
: QY 1 VRROLPEEENPAKRLFLLLTIVFC-----QILMAEGVPAPLPEDAPNAASLA--- 51
: | | | | | | | | | | | | | | | | | | | | | |
: Db 584 VNDNAPIEPR-----TIFFCERNPKQVINIHD---ADLPNTSPFTAE LTHGR 630
:
: QY 52 -----PTVPSPVLEPFLNTSEPDSYDALDLSLTF LQQH 82
: | | | | | | | | | | | | | | | | | | | | | |
: Db 631 VPNTIOYNDPTQESIILKP-KMALEVGDYKINLKLMDNQ 670
:
: RESULT 11
: PCT-US95-05518-2
: Sequence 2, Application PC/TUS9505518
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
: E-cadherin Interactions with T Lymphocytes
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wolf, Greenfield & Sacks P.C.
: STREET: 600 Atlantic Avenue
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/05518
: FILING DATE: herewith
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/237,919
: FILING DATE: 3 May 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Plumer, Elizabeth R.
: REGISTRATION NUMBER: 36,637
: REFERENCE/DOCKET NUMBER: B0801/7023
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-720-3500
: TELEFAX: 617-720-2441
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 878 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US95-05518-2
:
: Query Match 15.5% Score 68.5; DB 4; Length 878;
: Best Local Similarity 25.7%; Pred. No. 9;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:05:37 ; Search time 54.97 Seconds
(without alignments)
53.496 Million cell updates/sec

Title: US-08-799-910-10_COPY_71_156

Perfect score: 442

Sequence: 1 VRQLPVEENPAKRLFL.....EPDYLALDLSTFLQHPAAF 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36:**

1: /cgn2_2/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	442	100.0	156	18 W36006	Human Fchd605 gene
2	442	100.0	156	21 Y45017	Protein encoded by
3	79.5	18.0	1596	21 Y68821	Amino acid sequenc
4	71.5	16.2	610	15 R51116	Platelet glycoprot
5	71.5	16.2	610	15 R56664	Mutant platelet g
6	71.5	16.2	610	17 R89436	Mutated platelet g
7	71.5	16.2	610	18 W18201	Platelet glycoprot
8	69.5	15.7	739	18 W25790	Gene 036 product d
9	68.5	15.5	204	18 W14578	Streptococcus pneu
10	68.5	15.5	370	21 Y99428	Human PRO1431 (UNQ
11	68.5	15.5	878	15 R55060	Sequence of human
12	68.5	15.5	878	16 R85487	Human E-cadherin p

13	68.5	15.5	882	20 Y09375	Wild-type E-cadher
14	68.5	15.5	1337	16 R85203	huDEP-1. Homo sap
15	67	15.2	541	19 W37148	Mammalian Ena (Men
16	67	15.2	783	19 W37151	Mouse neural Mena+
17	67	15.2	787	19 W37152	Mouse neural Mena+
18	67	15.2	802	19 W37153	Mouse neural Mena+
19	66.5	15.0	131	18 W27646	Secreted protein A
20	66.5	15.0	131	18 W44082	Human secreted pro
21	65.5	14.8	128	18 W14577	Streptococcus pneu
22	65.5	14.8	286	17 W03566	Mycobacterium tube
23	65.5	14.8	325	17 W03565	Mycobacterium tube
24	65.5	14.8	332	18 W32418	Mycobacterium tube
25	65.5	14.8	332	18 W32350	Mycobacterium tube
26	65.5	14.8	332	19 W81683	M. tuberculosis im
27	65.5	14.8	332	19 W64322	Mycobacterium tube
28	65.5	14.8	332	20 Y39082	M. tuberculosis an
29	65.5	14.8	332	20 Y38945	M. tuberculosis re
30	65.5	14.8	456	20 Y17067	Human 3-OST-4 prot
31	65.5	14.8	652	20 Y39225	M. tuberculosis fu
32	65.5	14.8	652	20 Y39082	M. tuberculosis fu
33	65.5	14.8	802	19 W81746	M. tuberculosis fu
34	65.5	14.8	802	19 W64379	Mycobacterium anti
35	65.5	14.8	802	20 Y32063	Mycobacterium tube
36	65.5	14.8	802	20 Y39224	M. tuberculosis fu
37	65.5	14.8	802	20 Y39176	M. tuberculosis fu
38	65.5	14.8	802	20 Y39081	M. tuberculosis fu
39	65.5	14.8	802	20 Y39033	M. tuberculosis fu
40	65.5	14.8	1093	16 R66460	AF-17 protein. Ho
41	65	14.7	328	19 W36984	E. coli zipA prote
42	65	14.7	427	21 Y68748	Amino acid sequenc
43	64.5	14.6	147	15 R60799	Rape abscission/de
44	64.5	14.6	194	18 W14584	Streptococcus pneu
45	64.5	14.6	412	20 Y49151	Amino acid sequenc

ALIGNMENTS

RESULT 1

ID W36006 standard; Protein: 156 AA.

AC W36006;

DT 03-MAR-1998 (first entry)

DE Human Fchd605 gene product.

XX Fchd605 gene; differential expression; monocyte; human;
KW foam cell; cardiovascular disease; atherosclerosis; ischaemia;
KW reperfusion; hypertension; restenosis; arterial inflammation;
KW therapy; diagnosis; drug screening; marker.

OS Homo sapiens.

PN W09730065-A1.

PD 21-AUG-1997.

XX 14-FEB-1997; 97WO-US02291.

PR 13-FEB-1997; 97US-0799910.

XX 16-FEB-1996; 96US-0011787.

PA (MILL-) MILLENNIUM PHARM INC.

XX Falb DA;

DR WPI; 1997-424966/39.

XX N-PSDB; T94471.

XX New genes differentially expressed in cardiovascular disease - used

PT for diagnosis, drug screening and treatment of cardiovascular

CC three-dimensional coordinates obtained from the crystals, is useful
 CC for identifying an agent that stabilizes the Ras-Sos complex. The
 CC crystals are also useful for identifying agents that inhibit the
 CC formation of Ras-Sos complex. Ras and Sos fragments are useful for
 CC growing a crystal of a protein-ligand complex. Agents that stabilize
 CC or inhibit the formation of Ras-Sos complex are useful in the
 CC treatment of cancer. The present sequence represents a Drosophila
 CC Sos protein.

XX SQ Sequence 1596 AA;

Query Match 18.0%; Score 79.5; DB 21; Length 1596;
 Best Local Similarity 28.3%; Pred. No. 4.2;
 Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRRQLPVE-----EPNPAKRLFLLLTIVFCOILMAEGVPAPIPEDAPNAASLAPT 53
 DB 1423 lrnsaiekraaatsqpnqaagpistltvcsqavatdepilpisp-----aasestt 1477
 QY 54 --PVSPVLEPF--NLTSEP-----SDYALDLSTFLQV----HPAAF 86
 DB 1478 tsptpamspnshpvestssyahqlrm:qgqgqgthpaay 1523

RESULT 4

ID R51116 standard; protein; 610 AA.

AC R51116;

XX R51116;

DT 23-SEP-1994 (first entry)

DE Platelet glycoprotein Ib alpha.

KW Platelet; glycoprotein; imaging; thrombolytic agent;
 KW tissue plasminogen activator; tPA; pro-urokinase; urokinase;
 KW streptokinase; Bernard-Soulier disease; thrombus; aggregation;
 KW anisoylated plasminogen-streptokinase activator complex; adhesion;
 KW inhibition.

XX Homo sapiens.

XX US5298239-A.

XX 29-MAR-1994.

XX 07-OCT-1991; 91US-0770968.

XX 07-OCT-1991; 91US-0770968.

XX 15-JAN-1992; 92US-0821717.

XX (UYNV) UNIV NEW YORK STATE RES FOUND.

XX Cunningham D, Finch CN, Lyle VA, Miller JL;

XX WPI; 1994-100287/12.

XX Platelet glycoprotein Ib alpha with an amino acid substn at
 XX position 57 - has reduced reactivity with Von Willebrand factor,
 XX and can be used to inhibit platelet aggregation and inhibition

XX Claim 1; Columns 3-8; 20pp; English.

CC A substitution in platelet glycoprotein Ib alpha (Leucine 57 to
 CC phenylalanine) underlies a form of Bernard-Soulier disease. The
 CC mutated glycoprotein can be used in compositions to inhibit
 CC platelet aggregation/adhesion. The glycoprotein may be labelled and
 CC used as an imaging agent and may also be bound to a thrombolytic
 CC agent, preferably tissue plasminogen activator (tPA),
 CC (pro) urokinase, streptokinase, anisoylated plasminogen-streptokinase
 CC activator complex, tPA analogues or a protease, allowing localisation
 CC of the thrombolytic agent to a thrombus.

XX SQ Sequence 610 AA;

Query Match 16.2%; Score 71.5; DB 15; Length 610;
 Best Local Similarity 45.0%; Pred. No. 9.5;
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

QY 36 PAPLP-----PEDAPNAASLAPTVPSPVLEPFFNLTSPTS 69
 DB 366 ptpsttsepvppeapnmttleptpspttpep---tsepa 402

RESULT 5

ID R56664 standard; protein; 610 AA.

XX R56664;

XX 23-SEP-1994 (first entry)

DE Mutant platelet glycoprotein Ib alpha.

KW Platelet; glycoprotein; imaging; thrombolytic agent;
 KW tissue plasminogen activator; tPA; pro-urokinase; urokinase;
 KW streptokinase; Bernard-Soulier disease; thrombus; aggregation;
 KW anisoylated plasminogen-streptokinase activator complex; adhesion;
 KW inhibition.

XX Homo sapiens.

XX US5298239-A.

XX 29-MAR-1994.

XX 07-OCT-1991; 91US-0770968.

XX 07-OCT-1991; 91US-0770968.

XX 15-JAN-1992; 92US-0821717.

XX (UYNV) UNIV NEW YORK STATE RES FOUND.

XX Cunningham D, Finch CN, Lyle VA, Miller JL;

XX WPI; 1994-100287/12.

XX Platelet glycoprotein Ib alpha with an amino acid substn at
 XX position 57 - has reduced reactivity with Von Willebrand factor,
 XX and can be used to inhibit platelet aggregation and inhibition

XX Claim 1; 20pp; English.

CC A substitution in platelet glycoprotein Ib alpha (Leucine 57 to
 CC phenylalanine) underlies a form of Bernard-Soulier disease. The
 CC mutated glycoprotein can be used in compositions to inhibit
 CC platelet aggregation/adhesion. The glycoprotein may be labelled and
 CC used as an imaging agent and may also be bound to a thrombolytic
 CC agent, preferably tissue plasminogen activator (tPA),
 CC (pro) urokinase, streptokinase, anisoylated plasminogen-streptokinase
 CC activator complex, tPA analogues or a protease, allowing localisation
 CC of the thrombolytic agent to a thrombus.

XX SQ Sequence 610 AA;

Query Match 16.2%; Score 71.5; DB 15; Length 610;
 Best Local Similarity 45.0%; Pred. No. 9.5;
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

QY 36 PAPLP-----PEDAPNAASLAPTVPSPVLEPFFNLTSPTS 69
 DB 366 ptpsttsepvppeapnmttleptpspttpep---tsepa 402

QY 21 LTVFCQILMA-----EGVPAPLPPEDAPNAASLAPTPVSPVLEPFLNLTSEP-SDYAL 73.
 DB 793 lttvscgkmaaptrtcttgicdpppgpsnitsvshsvkxfgsfgeashgpkikayav 852
 QY 74 DLSTFLQHPAA 85
 DB 853 ltttgeaghpsa 864

RESULT 15
 W37148
 ID W37148 standard; Protein: 541 AA.
 AC W37148;
 DT 06-JUL-1998 (first entry)
 DE Mammalian Ena (Mena).
 KW Mena protein; mammalian Ena; Enabled; Evi protein; cytoskeleton;
 KW cell morphology; cell adhesion; cell differentiation; cell growth;
 KW cell motility; mouse.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT Domain 1..113
 FT Peptide /note= "Ena-VASP homology domain 1 (EVH1)"
 FT Peptide 156..160
 FT Peptide /note= "LERER repeat"
 FT Peptide 166..170
 FT Peptide /note= "LERER repeat"
 FT Peptide 171..175
 FT Peptide /note= "LERER repeat"
 FT Peptide 176..180
 FT Peptide /note= "LERER repeat"
 FT Peptide 199..203
 FT Peptide /note= "LERER repeat"
 FT Peptide 207..211
 FT Peptide /note= "LERER repeat"
 FT Modified-site 236
 FT Binding-site /note= "O-phosphorylated"
 FT Binding-site 305..310
 FT Binding-site /note= "mediates binding to profilin"
 FT Binding-site 316..321
 FT Binding-site /note= "mediates binding to profilin"
 PN W09801755-A1.
 PD 15-JAN-1998.
 PF 03-JUL-1997; 97WO-US11669.
 PR 05-JUL-1996; 96US-0675815.
 PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
 XX Gertler FB, Niebuhr K, Soriano P, Wehland J;
 WPI: 1998-101197/09.
 DR N-PSDB: V02996.
 XX Detection of modulators of Mena and Ena-VASP-like genes and proteins
 PT - used in control of cytoskeletal dynamic events in normal and
 PT abnormal cell morphology, adhesion, motility, growth and
 PT differentiation
 XX
 PS Example 1: Page 54-56; 77pp; English.
 CC This polypeptide comprises murine Mena (mammalian Ena) that shows
 CC significant amino acid similarity to Drosophila Ena and which
 CC exhibits a broad pattern of tissue distribution in neurons,

CC fibroblasts, kidney epithelium, muscle, neural crest and
 CC hematopoietic cells. Its amino acid sequence was deduced from
 CC a cDNA clone (see V02996) obtained from a mouse embryonic stem cell
 CC cDNA library. 2 Novel mammalian genes, Mena and Ena-VASP-like
 CC (Evl, see V02997), encoding novel proteins Mena and Evl (see
 CC W37149) are disclosed. Mena and Evl proteins have a discrete, EVH1
 CC functional domain responsible for Mena binding to Listeria, and to
 CC the cytoskeletal proteins zyxin and vinculin. The EVH1 domain of
 CC Mena is also responsible and sufficient for targeting localisation
 CC of Mena and Mena-based fusion proteins to focal adhesions, and to
 CC the surface of Listeria cells at the polar site of induction of
 CC actin comet tail formation. Based on the disclosed Mena and Evl
 CC genes and proteins, a variety of methods and compositions are
 CC provided for screening, isolating and characterising endogenous and
 CC exogenous factors, drugs and therapeutic agents useful to evaluate
 CC and/or control cytoskeletal dynamic events involved in normal and
 CC abnormal cell morphology, adhesion, motility, growth and/or
 CC differentiation. A method of detecting a modulator of Mena
 CC activity/expression is claimed.

XX Sequence 541 AA;
 SQ

Query Match 15.2%; Score 67; DB 19; Length 541;
 Best Local Similarity 29.9%; Pred. NO. 25;
 Matches 20; Conservative 4; Mismatches 23; Indels 20; Gaps 2;

QY 6 PVEEPNPAKRLFLFLLLTVFCQILMAEGVPAPLPPEDAPN-----AASLAPTPVSPVLEP 61
 DB 268 paesptpggvl-----gppapppppplpsgpayasalppppppppp 311
 QY 62 FNLTSEP 68
 DB 312 lpstgpp 318

Search completed: January 30, 2001, 22:05:37
 Job time: 11886 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:04:55 ; Search time 64.84 Seconds
(without alignments)
281.993 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPSDYALDLSTFLOQHPAAF 156

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_15.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	771.5	93.7	193	4	09Y4F5
2	122.5	14.9	423	4	09YMX0
3	114.5	13.9	1668	4	09Y519
4	114.5	13.9	2971	4	09Y519
5	114	13.9	3247	12	065553
6	112.5	13.7	649	5	09V884
7	110.5	13.4	1011	3	09P944
8	109.5	13.3	552	10	09G343
9	109.5	13.3	1315	10	09SPM0
10	109	13.2	285	3	P78977
11	109	13.2	1709	4	015047
12	109	13.2	3325	12	09LBT9
13	108.5	13.2	233	2	09L568
14	107.5	13.1	203	2	09RCX9
15	107.5	13.1	581	5	020517
16	107.5	13.1	1013	5	09VY77
17	107	13.0	426	2	09LAY5
18	107	13.0	496	12	098457
19	106.5	12.9	901	2	044562

20	106.5	12.9	1498	4	09Y4F5
21	106	12.9	1029	12	09YMX0
22	106	12.9	2066	12	P89920
23	105.5	12.8	249	2	09L570
24	105	12.8	296	5	09V729
25	105	12.8	351	10	039492
26	105	12.8	549	12	089370
27	104.5	12.7	243	2	09L567
28	104.5	12.7	244	2	09L565
29	104.5	12.7	535	2	P74745
30	104.5	12.7	616	12	096716
31	104	12.6	395	2	09LAY2
32	104	12.6	408	2	09LAY0
33	103	12.5	241	5	045522
34	102.5	12.5	288	12	084565
35	102.5	12.5	767	5	062537
36	102	12.4	236	2	09L569
37	102	12.4	331	5	09V728
38	102	12.4	708	10	09SX31
39	101.5	12.3	344	11	089037
40	101	12.3	339	12	084465
41	101	12.3	395	2	09LAZ1
42	101	12.3	437	2	09LAY4
43	101	12.3	739	2	09ROT4
44	101	12.3	852	4	09NP71
45	100	12.2	802	11	P70433

ALIGNMENTS

RESULT 1

075353 ID O75353 PRELIMINARY; PRT; 193 AA.
AC O75353;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ANTI-DEATH PROTEIN.
GN IEX-1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98369175; PubMed=9703517;
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival.";
RL Science 281:998-1001(1998).
DR EMBL; AF039067; AAC32558.1;
DR EMBL; AF071596; AAC72344.1;
SQ SEQUENCE 193 AA; 7927D9D3FFBC7C57 CRC64;

Query Match 93.7%; Score 771.5; DB 4; Length 193;
Best Local Similarity 78.8%; Pred. No. 8.1e-64;
Matches 152; Conservative 1; Mismatches 3; Indels 37; Gaps 1;
QY 1 MCHSRCHPTMTILOAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASGRHK 60
DB 1 MCHSRCHPTMTILOAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASGRHK 60
QY 61 RSRRLVYPR-----VVRQLPVEEPNPA 83
DB 61 RSRRLVYPR-----VVRQLPVEEPNPA 120
QY 84 KRLLFLLLTIVFCOILMAEGVPAPLPEDAPNAASLAPTPVSPVLPEFNLTSPEPSYAL 143
DB 121 KRLLFLLLTIVFCOILMAEGVPAPLPEDAPNAASLAPTPVSPVLPEFNLTSPEPSYAL 180
QY 144 DLSTFLQOHPAAF 156

DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE UL36.
GN UL36.
OS Bovine herpesvirus 1
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwyz M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
RA Thilly E., Paces V.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=COOPER;
RA Schwyz M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA Laboisserie S., Misra V., Vlcek C., Paces V.;
RL Vet. Microbiol. 0:0-0(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwyz M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=JURA;
RX MEDLINE=97164286; PubMed=9010999;
RA Schwyz M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA Laboisserie S., Misra V., Vlcek C., Paces V.;
RT "Gene contents in a 31-kb segment at the left genome end of bovine
herpesvirus-1".
RL Vet. Microbiol. 53:67-77(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwyz M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwyz M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z78205; CAB01605.1;
DR EMBL: AJ004801; CA006097.1;
SQ SEQUENCE 3247 AA; 332190 MW; 3AEAA72F8F001F6A CRC64;

Query Match 13.9%; Score 114; DB 12; Length 3247;
Best Local Similarity 31.18; Pred. No. 0.045;
Matches 41; Conservative 12; Mismatches 49; Indels 30; Gaps 7;

QY 9 PTMTIQAFTAPSTIPGPRRGSGPEIFTFDPLEPAAPAGPSASGRHKRRRLVLP 68
DB 2705 PITVLVAPVAPAPAPAPAPAPAPAPAPAPPPPP---APGAMGALSATRRPTRA--- 2759
QY 69 RVVRQLPVEENPAKRLFLILLITVFCQILMAEGVPAP----LPPEDAPNAASLAPT 124
DB 2760 -GARKSLPAAOPR--ORLL-----RSRSPASVPAPGSELVPP---PSGGALGSP 2803
QY 125 V-----SPVLEP 131
DB 2804 SFVPSRPPSLEP 2815

RESULT 6
ID Q9V884 PRELIMINARY; PRT; 649 AA.
AC Q9V884;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE CG18635 PROTEIN.

GN CG18635.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arlil J.F., Agbayani A., An H.-J., Andrews-Prankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brookslein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doip L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Fertaz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Hock J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster".
RL Science 287:2185-2195(2000).
DR EMBL: AE003802; AAF57791.1;
DR FLYBASE; FBgn0034279; CG18635.
SQ SEQUENCE 649 AA; 74414 MW; 6471AC52B49F9AB9 CRC64;

Query Match 13.7%; Score 112.5; DB 5; Length 649;
Best Local Similarity 25.8%; Pred. No. 0.013;
Matches 46; Conservative 26; Mismatches 67; Indels 39; Gaps 7;

QY 10 TMTILQA--PTPAPSTIPGPRRGSGPEIFTFDPLEPAAPAGPSASGRHKRRRLVLY 67
DB 342 TIEVLDSIPTAP--PPPKTAPPELVKPTTEEDPTPDLRLTSDNEHKRRRVR 398
QY 68 P-----RVVRQLPVEE--PNPAKRLFLLL-----TIVFCQILMAEGVPAP 108
DB 399 PLSQLETFKDWLLVNAQLYIKENVPHPKLLIGRLKDEDPEHT-----VLIPQSPKPS 453
QY 109 LPEDAPNAASLAPTPVSPVLEPNLTSEPSDY-----ALDLSTFLOQHPAF 156
DB 454 LTPKDVVDYFPRPTKIIVNGIQODDFAGRTVSFFGIAFITMRMLSLSTFCVYKRAF 511

RESULT 7
ID Q9P944 PRELIMINARY; PRT; 1011 AA.

RX MEDLINE-97127825; PubMed-8972576;
 RA Ramon A., Gil R., Burgal M., Sentandreu R., Valentin E.;
 RT "A novel cell wall protein specific to the mycelial form of *Yarrowia*
 RT *lipolytica*.";
 RL Yeast 12:1535-1548(1996).
 DR EMBL: Z81006; CAB02634.1;
 KW Signal.
 FT SIGNAL. 1 16 POTENTIAL.
 FT CHAIN 17 285 POTENTIAL.
 SQ SEQUENCE 285 AA; 30036 MW; OBA7FF1C515CD57D CRC64;

Query Match 13.2%; Score 109; DB 3; Length 285;
 Best Local Similarity 28.4%; Pred. No. 0.012;
 Matches 40; Conservative 11; Mismatches 62; Indels 28; Gaps 6;

QY 1 MCHSRCHPTMTILQATPAP-STIPGPRGSGPEIFTDPLPEAAAPAGPSASRGHR 59
 DB 122 ICTKTVLSTVLTPTVPVPTTTPAVEPKPTPEVPEVKPEPTP-EVPGVKPETRGPP 180
 QY 60 KRSRRVLYPRVVRNOLP-VEEPNPAKRLLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAA 118
 DB 181 AKPEPEVEPKPTPEVPEVRP-----EPTPAPLPFR--PSLR 218
 QY 119 SLAPTPVSPVLEFFNLTSPPS 139
 DB 219 SLRS---SPSLPPLPLPPSPS 236

RESULT 11
 OI5047 PRELIMINARY; PRT; 1709 AA.
 ID OI5047;
 AC OI5047;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE KIRA0339 PROTEIN (FRAGMENT).
 GN KIRA0339.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 DR EMBL: AB002337; BAA20797.2;
 DR INTERPRO: IPR000504;
 DR INTERPRO: IPR001214;
 DR INTERPRO: IPR002965;
 DR PFAM: PF00076; Trm. 1;
 DR PFAM: PF00856; SET. 1;
 DR PRINTS: PR01217; PRICHEXTNSN.
 FT NON_TER 1
 SQ SEQUENCE 1709 AA; 186231 MW; DCF2E0FF716B672C CRC64;

Query Match 13.2%; Score 109; DB 4; Length 1709;
 Best Local Similarity 32.6%; Pred. No. 0.069;
 Matches 44; Conservative 10; Mismatches 53; Indels 28; Gaps 7;

QY 6 SCHPTMTILQATPAPSTIPGPRGSGPEIFTDPLPEAAAPAGPSASRGHRKRRV 65
 DB 1078 SASPPPEVPVPTAPVEVPVPERVAGSPV---TLPQEASPA-RVAGTTEESPPSAPL 1133
 QY 66 LYPRVVRQLPVEEPNPAKR-----LLFL-----LTIVFCQILMAEGVPAPLP 111
 DB 1134 RPP-----EPPAGPPAPAPRDERPSSPIPLPPPKRRKRTVPSAI-----EVVPAPEPP 1184

QY 112 EDAPNAASLAPTPVS 126
 DB 1185 PATPPQAKF-PCPAS 1198

RESULT 12

QY 091BT9 PRELIMINARY; PRT; 3325 AA.
 AC QY1BT9;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE MAJOR TEGUMENT PROTEIN.
 GN UL36;
 OS Turkey herpesvirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
 OX NCBI_TaxID=10390;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GA;
 RX MEDLINE=92237304; PubMed=1315048;
 RA Vares D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;
 RT "Varek disease virus encodes a basic-leucine zipper gene resembling
 RT the fos/jun oncogenes that is highly expressed in lymphoblastoid
 RT tumors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GA;
 RA Lee L.F., Wu P., Sui D., Ren D., Kung H.J.;
 RT "The Complete UL Sequence of Serotype I Marek's Disease Virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).
 DR EMBL: AF147806; AAF66771.1;
 SQ SEQUENCE 3325 AA; 365698 MW; 5D4A84EF719B99FD CRC64;

Query Match 13.2%; Score 109; DB 12; Length 3325;
 Best Local Similarity 25.2%; Pred. No. 0.13;
 Matches 31; Conservative 14; Mismatches 52; Indels 26; Gaps 3;
 QY 16 APTPAPSTIPGPRGSGPEIFTDPLPEAAAPAGPSASRGHRKRRVLYPRVVRQL 75
 DB 2802 SPAPKSPAPKPPPPDPD---FKPSPAPKSPASKPSASKPSASKPPPPADSKPS 2858
 QY 76 PVSEPNKAKRLFLLLLTIVFCQILMAEGVPAPLPDPEDAPNAASLAPTP-----VSPV 128
 DB 2859 PAPKPPPP-----TPDSKPSAPKPSASKPSASKPLPVPPFPNSDSKTSVP 2902
 QY 129 LEP 131
 DB 2903 PNP 2905

RESULT 13

QY 091568 PRELIMINARY; PRT; 233 AA.
 ID QY1568;
 AC QY1568;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE PSPA (FRAGMENT).
 GN PSPA.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=39;
 RA Beall B.W.;
 RT "PSPA sequence types from multiresistant pneumococci.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 21:28:35 ; Search time 21.17 seconds
(without alignments)
500.354 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823
Sequence: 1 MCHSRCHPTMTILQAPTPA.....EPDYLDTFLQHPAAF 156

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*

1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	819	99.5	156	2 JC5537	differentiation-de
2	525	63.8	153	2 S33363	gly96 protein - mo
3	112.5	13.7	283	2 S13383	hydroxyproline-ric
4	109.5	13.3	552	2 T08148	proline-rich myros
5	107.5	13.1	581	2 T22341	hypothetical prote
6	107	13.0	496	2 T17908	proline/lysine-ric
7	106.5	12.9	901	2 A49227	siatidase - Actino
8	106.5	12.9	1323	2 T00037	hypothetical prote
9	106	12.9	1029	2 T30351	mucin-like protein
10	105	12.8	351	2 S50754	hypothetical prote
11	105	12.8	549	2 T17525	proline-rich prote
12	105	12.8	3149	1 Q76E8	BPLF1 protein - hu
13	104.5	12.7	535	1 S76953	protein kinase (BC
14	104	12.6	815	2 B56708	extracellular sign
15	103	12.5	241	2 T22216	hypothetical prote
16	102.5	12.5	288	2 T17737	proline-rich prote
17	102.5	12.5	3866	2 B48205	All-1 protein -GTE
18	102.5	12.5	3869	2 A48205	All-1 protein +GTE
19	101	12.3	339	2 T17636	proline-rich prote
20	100	12.2	620	2 S06733	hydroxyproline-ric
21	99.5	12.1	279	2 T10361	hypothetical prote
22	98.5	12.0	3164	1 WMBEH6	UL36 protein - hum
23	98	11.9	473	2 S50755	hypothetical prote
24	97.5	11.8	225	2 T17815	proline-rich prote
25	97.5	11.8	2157	2 S71461	proline-rich prote
26	97.5	11.8	3338	2 T42761	Bassoon protein -
27	97	11.8	461	2 T10741	extensin-like prot
28	96.5	11.7	168	2 S52994	arabinogalactan-li
29	96	11.7	185	2 T00519	proline-rich prote

30 95.5 11.6 524 2 A75588 probable protein k
31 95.5 11.6 537 2 A46611 myosin-binding pro
32 95.5 11.6 753 2 JQ0532 OP protein - Kanne
33 95.5 11.6 1596 2 A41216 guanine nucleotide
34 95 11.5 202 2 JQ096A hydroxyproline-ric
35 95 11.5 377 2 A48018 mucin 7 precursor,
36 95 11.5 416 1 SKXLAG dermal gland prote
37 95 11.5 801 2 T29018 hypothetical prote
38 95 11.5 3511 2 A59295 unconventional myo
39 94.5 11.5 510 1 S43692 transcription fact
40 94.5 11.5 758 2 A29253 finger protein hun
41 94.5 11.5 2342 2 T13412 hypothetical prote
42 94 11.4 328 2 JQ0985 hydroxyproline-ric
43 94 11.4 925 2 S27920 nuclear antigen EB
44 94 11.4 1290 2 T00018 period protein hom
45 93.5 11.4 316 2 T31880 hypothetical prote

ALIGNMENTS

RESULT 1

JC5537
differentiation-dependent protein DIF-2 - human
C:Species: Homo sapiens (man)
C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 07-May-1999
C:Accession: JC5537
R:Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.
Biochem. Biophys. Res. Commun. 235, 4-9; 1997
A:Title: Identification and characterization of a novel monocyte/macrophage differ
A:Reference number: JC5537; MUID:97339426
A:Accession: JC5537
A:Molecule type: mRNA
A:Residues: 1-156 <PIF>
A:Experimental source: monocyte
A>Note: the authors translated the codon CCG for residue 106 as Arg

Query Match 99.5% Score 819; DB 2; Length 156;
Best Local Similarity 99.4% Pred. No. 1.2e-58;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRGSGPEITFDPLPEPAAAPAGRPSASGRHK 60
|||||
DB 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRGSGPEITFDPLPEPAAAPAGRPSASGRHK 60
QY 61 RRRVLYPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAERGVPAPLPEDAPNAASL 120
|||||
DB 61 RRRVLYPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAERGVPAPLPEDAPNAASL 120
QY 121 APTVPSPVLEPNLTSEPSDYALDLSTFLQHPAAF 156
|||||
DB 121 APTVPSPVLEPNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 2

S33363
gly96 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C:Accession: S33363
R:Charles, C.H.; Yoon, J.K.; Simske, J.S.; Lau, L.F.
Oncogene 8, 797-801, 1993
A:Title: Genomic structure, cDNA sequence, and expression of gly96, a growth factor.
A:Reference number: S33363; MUID:93173526
A:Accession: S33363
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-153 <CHA>
A:Cross-references: EMBL:X67644
C:Genetics:
A:Introns: 70/3
C:Keywords: transmembrane protein

Result No.	Query			ID	Description
	Score	Match	Length		
1	823	100.0	156	1	IEX1_HUMAN
2	525	63.8	153	1	IEX1_MOUSE
3	112.5	13.7	283	1	EXTN_SORBI
4	105	12.8	3149	1	TEGU_EBV
5	104	12.6	815	1	MK07_HUMAN
6	102.5	12.5	381	1	MODD_MYCAV
7	102.5	12.5	3866	1	HRX_MOUSE
8	101.5	12.3	268	1	NO20_MEDTR
9	100	12.2	620	1	EXTN_TOBAC
10	99.5	12.1	279	1	Y091_NPVOP
11	98.5	12.0	3164	1	TEGU_HSV11
12	95.5	11.6	537	1	MYPH_CHICK
13	95	11.5	439	1	XP2_XENLA
14	94.5	11.5	510	1	ERM_HUMAN
15	94.5	11.5	758	1	HUNB_DROME
16	94.5	11.5	1595	1	SOS_DROME
17	94	11.4	1290	1	PER1_HUMAN
18	93.5	11.4	1402	1	IF4G_RABIT
19	93	11.3	817	1	VRP1_YEAST
20	92.5	11.2	605	1	BRL1_EBV
21	92.5	11.2	625	1	NIFA_AZOBR
22	92	11.2	534	1	APG_AZATH
23	92	11.2	3669	1	HRX_HUMAN
24	91.5	11.1	267	1	EXTN_WA12E
25	91.5	11.1	615	1	MUTL1_ECOLI
26	91.5	11.1	1133	1	SREL1_CRIGR
27	90	10.9	316	1	CDNC_HUMAN
28	90	10.9	449	1	APG_BRANA
29	89.5	10.9	1487	1	ICP4_HSVK
30	89.5	10.9	2715	1	TRX2_HUMAN
31	89	10.8	1291	1	PER1_MOUSE
32	88.5	10.8	865	1	CPN_DROME
33	88	10.7	4774	1	VTF3_TTV1V

DR PROSITE: PS50014; BROMODOMAIN.2; 1;
 KW Proto-oncogene: Chromosomal translocation; DNA-binding;
 KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
 KW Alternative splicing; Polymorphism.

FT NON_TER 1
 FT DNA_BIND 67 78 A.T HOOK (BY SIMILARITY).
 FT DNA_BIND 115 125 A.T HOOK (BY SIMILARITY).
 FT DNA_BIND 199 207 A.T HOOK (BY SIMILARITY).
 FT DOMAIN 1330 1381 PHD 1.
 FT DOMAIN 1383 1432 PHD 2.
 FT DOMAIN 1465 1529 PHD 3.
 FT DOMAIN 1605 1650 BROMODOMAIN (DIVERGENT).
 FT DOMAIN 3737 3846 SET DOMAIN.
 FT DOMAIN 35 41 POLY-GLY.
 FT DOMAIN 459 469 PRO-RICH.
 FT DOMAIN 1231 1238 POLY-PRO.
 FT DOMAIN 3533 3536 POLY-GLU.
 FT DOMAIN 3693 3697 POLY-GLU.
 FT VARSPLIC 1503 1505 MISSING (IN ISOFORM 2).
 FT VARIANT 1497 1497 K -> T.
 SQ SEQUENCE 3866 AA; 420976 MW; ADFC55E14E806F1D CRC64;

Query Match 12.5%; Score 102.5; DB 1; Length 3866;

Best Local Similarity 26.8%; Pred. No. 1.8;

Matches 41; Conservative 19; Mismatches 48; Indels 45; Gaps 8;

QY 12 TILQAP-TPASTIPGRSGPEIFTDPLPE-----PAAAPAGRPS--ASR 56
 DB 1129 TSVKSPLEPAQAARPPREPAPKSSSEPPRPKPKVEKSEGGAPAPAPPEPKQVAP 1180
 QY 57 GHKRSRRVLYPRVRRQLPVEPNPAKRLLELLTIVFCQILMAEGVPA-----PLP 110
 DB 1189 ASRKSSKQVSQAAV---VPPQPPSTAP-----OKKEAPKAVSEPKKKQPPP 1233
 QY 111 PEDAPNAA---SLAPTSPVLEPNLTSEPSD 140
 DB 1234 PEPGPEQSKQKVAPLPSPVKQ-----KPKD 1260

RESULT 8
 ID NO20_MEDTR STANDARD; PRT: 268 AA.
 AC N9020_MEDTR
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE EARLY NODULIN 20 PRECURSOR (N-20).
 GN ENO20.
 OS Medicago truncatula (Barrel medic).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Medicago.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. JEMALONG;
 RX MEDLINE=96187258; PubMed=9526510;
 RA Greene E.A., Erard M., Dedieu A., Barker D.G.;
 RT "Mcnod16 and 20 are members of a family of phytoeyanin-related early nodulins".
 RL Plant Mol. Biol. 36:775-783(1998).
 CC -!- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC EMBL: X99467; CAA67830.1; -
 DR Nodulation; Signal.

FT SIGNAL 1 22 BY SIMILARITY.
 FT CHAIN 23 268 EARLY NODULIN 20.
 FT DOMAIN 23 7 PLASTOCYANIN-LIKE.
 FT DOMAIN 136 145 POLY-PRO.
 SQ SEQUENCE 268 AA; 28668 MW; 6FA29C5798C75F91 CRC64;
 Query Match 12.3%; Score 101.5; DB 1; Length 268;
 Best Local Similarity 28.3%; Pred. No. 0.15;
 Matches 41; Conservative 12; Mismatches 31; Indels 61; Gaps 9;
 QY 13 ILQAPTAPS-----TIPGRSGPEIFTDPLPEPAAAPAGRPSASGRHKRSRR 64
 DB 132 VLSPPPPPTPRSSSTPIPHPPRSPLS---PPSPSPSPSPSPSPSPSPSPSPS 178
 QY 65 VLYPRVRRQLPVEPNPAKRLLELLTIVFCQILMAEGVPAIP-----PEDAPNA 117
 DB 179 -----PRSTPI--PHPRKR-----SPASPSPSLSKSPSPSPSPSPSP 211
 QY 118 ASLAPTVPSPV--LEPFLNLTSEPSD 140
 DB 212 -SLAPSPSDSVASLAP---SSSPSD 232

RESULT 9

EXTN_TOBAC

ID EXTN_TOBAC STANDARD; PRT: 620 AA.
 AC P13983;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-1992 (Rel. 21, Last annotation update)
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HRCPT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 OC Solanales; Solanaceae; Nicotiana.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. XANTHI; TISSUE=LEAF;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation".
 RL Genes Dev. 3:1639-1646(1989).
 CC -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE MAIN ROOT.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN GLYCOSYLATED.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC EMBL: X13885; CAA32090.1; -
 DR PIR: S06733; S06733.
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 ?
 FT CHAIN ? 620 EXTENSIN.
 FT REPEAT 70 73 H-A-P-P.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 1.
 FT REPEAT 236 260 2.
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.

RT Molecular cloning of chicken myosin-binding protein (MyBP) H (86-kDa
RT protein) reveals extensive homology with MyBP-C (C-protein) with
RL conserved immunoglobulin C2 and fibronectin type III motifs.";
RL J. Biol. Chem. 268:3670-3676(1993).
CC -!- FUNCTION: BINDS TO MYOSIN; PROBABLY INVOLVED IN INTERACTION WITH
CC -!- THICK MYOFILAMENTS IN THE A-BAND.
CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE. SEEMS TO BE ALSO EXPRESSED IN
CC THE SLOW TONIC ALD MUSCLE. NOT DETECTED IN GIZZARD OR HEART.
CC -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE MYBP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L05605; AAA21418.1;
DR INTERPRO: IPR001777;
DR INTERPRO: IPR002965;
DR INTERPRO: IPR003006;
DR PFAM: PF00041; fn3; 2.
DR PFAM: PF00047; 19; 2.
DR PRINTS: PR00014; FNTYPEIII.
DR PRINTS: PR01217; PRICHEXTENSN.
KW Immunoglobulin domain; Cell adhesion; Muscle protein; Thick filament;
KW Repeat.
FT DOMAIN 135 221 FIBRONECTIN TYPE-III.
FT DOMAIN 253 312 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 331 416 FIBRONECTIN TYPE-III.
FT DOMAIN 458 518 IG-LIKE C2-TYPE DOMAIN.
FT CONFLICT 2 9 T -> G (IN AA SEQUENCE).
FT CONFLICT 9 2 A -> P (IN AA SEQUENCE).
FT CONFLICT 15 15 A -> K (IN AA SEQUENCE).
SQ SEQUENCE 537 AA; 58678 MW; 08C4CF0FE1DD233 CRC64;

Query Match 11.6%; Score 95.5; DB 1; Length 537;
Best Local Similarity 27.5%; Pred. No. 0.89;
Matches 38; Conservative 6; Mismatches 61; Indels 33; Gaps 5;
QY 9 PTMTLQATTPAPSTIPGRGSGPEIFTDFLPPEAAAPAGRPASRGHRKSRRLVLP 68
DB 38 PTPKSGHAPTPEEHAPPKKEHAP-----PPKECHAPAPAAETPPAEPHPDAEQPAAP 92
QY 69 RVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPDPADPNAASLAPTPVSPV 128
DB 93 AAHAAPTTHAAPAH-----EEG-PPAAPADAP-----APEP----- 125
QY 129 LEFNLTSRPSDYALDLS 146
DB 126 -EPEKPEEPSVPLSLA 142

RESULT 13
XP2_XENLA STANDARD; PRT; 439 AA.
AC P17437; Q08944;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEG PROTEIN).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Piploidea;
OC Xenopodinae; Xenopus.
RN [1]
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A.
RX MEDLINE=92332564; PubMed=1629230;
RA Hauser F., Roeben C., Hoffmann W.;

RT "xp2, a new member of the P-domain peptide family of potential growth
RT factors, is synthesized in Xenopus laevis skin.";
RL J. Biol. Chem. 267:14451-14455(1992).
RN [2]
RP SEQUENCE OF 3-439 FROM N.A.
RT TISSUE-SKIN:
RX MEDLINE=90127399; PubMed=22982993;
RA Gnachl M., Berger H., Thalhammer J., Kreil G.;
RT "Dermal glands of Xenopus laevis contain a polypeptide with a highly
RT repetitive amino acid sequence.";
RL FEBS Lett. 260:145-148(1990).
CC -!- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE
CC EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS
CC AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS OF THE PROTEINS ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE APEG FORM INCLUDES A LARGE
CC DOMAIN WITH 33 SEQUENCE REPEATS.
CC -!- TISSUE SPECIFICITY: SKIN.
CC -!- SIMILARITY: CONTAINS 2 P-TYPE (TREFOL) DOMAINS.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392
CC ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M90095; AAA50001.1;
DR EMBL: X51394; CAA35759.1; ALT_FRAME.
DR PIR: S07498; SKXLAG.
DR PIR: A37331; A37331.
DR HSP: P04155; IFS2.
DR INTERPRO: IPR000519;
DR PFAM: PF00088; TREFOL; 2.
DR PRINTS: PR00680; PTREFOL.
DR PROSITE: PS00025; P_TREFOL; 2.
KW Signal; Growth factor; Alternative splicing; Amphibian skin; Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.
FT MOD_RES 23 23 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT DOMAIN 26 343 33 X REPEATS OF G-[GE]-[AP](2,4)-A-E.
FT DOMAIN 350 391 P-TYPE 1.
FT DOMAIN 397 438 P-TYPE 2.
FT DISULFID 351 377 BY SIMILARITY.
FT DISULFID 361 376 BY SIMILARITY.
FT DISULFID 371 388 BY SIMILARITY.
FT DISULFID 398 424 BY SIMILARITY.
FT DISULFID 408 423 BY SIMILARITY.
FT DISULFID 418 435 BY SIMILARITY.
FT VARSPIC 26 343 MISSING (IN ISOFORM XP2).
FT CONFLICT 3 3 H -> S (IN REF. 2).
FT CONFLICT 18 18 C -> W (IN REF. 2).
SQ SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;

Query Match 11.5%; Score 95; DB 1; Length 439;
Best Local Similarity 25.7%; Pred. No. 0.8;
Matches 35; Conservative 6; Mismatches 57; Indels 38; Gaps 5;
QY 15 QAPTPAP-----STIPGRGSGPEIFTDFLPPEAAAPAGRPASRGHRKSRRLVLPV 70
DB 247 EAPAPAEAGEAPAPAEAGEAPA-----PAPAEAGEAPAPAEAGEAPAEAGEAPAP 301
QY 71 VRROLPE-----EPNPAKRLFLLLTIVFCQILMAEGVPAPLPDPADPNAASLAPTP 124
DB 302 AEGEAPAPAEAGGAPSP-----AEGGAPAAAPAEAGGAPAPAPVE 343
QY 125 VSPVLE-----PFLNT 135

FT	2N_FING	733	757	C2H2-TYPE.
FT	VARIANT	437	A -> P.	
FT	VARIANT	649	V -> M.	
SQ	SEQUENCE	758 AA;	AD74802EB856ACD7 CRC64;	

Query Match	11.5%	Score	94.5;	DB	1:	Length	758;
Best Local Similarity	24.6%;	Pred. NO.	1.5;	56;	Indels	57;	Gaps
Matches	43;	Conservative	19;	Mismatches			

Qy	2	CHRSCH-	-----PTWIILOAETPAPS----	TTPGRRSGSGPEITFDPLPEPAAPAGR	51
			::: ::		
			::: ::		
			::: ::		
Db	337	CHSEKHLRKYGHKPGMVLDEDGTNPNSLVIDVYTRRG	-----PKSKNGG	382	
Qy	52	PSASGRHKSRRLVPRVVRQLPVVEPNPAKRLLFLTLTIIVFCOILMAEGVP	-----	106	
Db	383	PIASGGSGSGRSKNVAAPAQQ	-OOOPAQ-----PVATSQLSAALQQGFPLVQGN	432	
Qy	107	-----APLPEDAPNAASLAAPTVPSPVLEFPNLTSPESDYALDLSTFLOQH	152		
Db	433	SAPPAASGPLVPLPASPAKSVASVEOTSLP	-----SPANLLPPLASLLOON	478	

Search completed: January 30, 2001, 23:12:24
Job time: 4005 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:12:24 ; Search time 43.61 seconds
(without alignments)
63.685 Million cell updates/sec

Title: US-08-799-910-10_COPY_71_156

Perfect score: 442

Sequence: 1 VRRQLVPEEPNPAKRLFL.....EPSDYALDLSLTLQHPAAF 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	442	100.0	156	1	1	1	442
2	273	61.8	153	1	1	1	273
3	78.5	17.8	1595	1	1	1	78.5
4	73.5	16.6	272	1	1	1	73.5
5	72.5	16.4	272	1	1	1	72.5
6	71.5	16.2	626	1	1	1	71.5
7	70.5	16.0	605	1	1	1	70.5
8	70.5	16.0	2715	1	1	1	70.5
9	69.5	15.7	1894	1	1	1	69.5
10	69	15.6	1402	1	1	1	69
11	68.5	15.5	181	1	1	1	68.5
12	68.5	15.5	882	1	1	1	68.5
13	68.5	15.5	1337	1	1	1	68.5
14	68.5	15.5	1844	1	1	1	68.5
15	68	15.4	1395	1	1	1	68
16	67	15.2	283	1	1	1	67
17	67	15.2	586	1	1	1	67
18	67	15.2	4289	1	1	1	67
19	66.5	15.0	3149	1	1	1	66.5
20	66	14.9	621	1	1	1	66
21	65.5	14.8	325	1	1	1	65.5
22	65.5	14.8	325	1	1	1	65.5
23	65.5	14.8	329	1	1	1	65.5
24	65.5	14.8	812	1	1	1	65.5
25	65.5	14.8	1093	1	1	1	65.5
26	65	14.7	209	1	1	1	65
27	65	14.7	328	1	1	1	65
28	65	14.7	427	1	1	1	65
29	65	14.7	724	1	1	1	65
30	64.5	14.6	686	1	1	1	64.5
31	64.5	14.6	1183	1	1	1	64.5
32	64	14.5	510	1	1	1	64
33	64	14.5	510	1	1	1	64

34	64	14.5	537	1	PTN5_HUMAN
35	64	14.5	541	1	PTN5_MOUSE
36	64	14.5	590	1	FTZ1_RHIME
37	64	14.5	966	1	SSN6_YEAST
38	63.5	14.4	443	1	HXA3_HUMAN
39	63.5	14.4	540	1	NUSA_MYCPN
40	63.5	14.4	816	1	QALF_NEUCR
41	63.5	14.4	1776	1	POLR_OYMW
42	63.5	14.4	1844	1	POLR_TYMCV
43	63	14.3	323	1	PF27_MOUSE
44	63	14.3	420	1	EFLA_HALWA
45	63	14.3	598	1	NR41_HUMAN

ALIGNMENTS

RESULT 1					
ID	1	STANDARD	PRT	156 AA	
AC	P46695; Q93044; Q92691;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY				
DE	PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN) (PACAP-RESPONSIVE				
DE	GENE 1 PROTEIN) (PRG1 PROTEIN) (DIFFERENTIATION-DEPENDENT GENE 2				
DE	PROTEIN) (DIF-2 PROTEIN).				
GN	IEX3 OR IEX1 OR PRG1 OR DIF2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=PLACENTA;				
RA	MEDLINE=96181295; PubMed=8603392;				
RA	Kondratyev A.D., Chung K.-N., Jung M.O.;				
RT	"Identification and characterization of a radiation-inducible				
RT	glycosylated human early-response gene.";				
RL	Cancer Res. 56:1498-1502(1996).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=96221139; PubMed=8653710;				
RA	Schaefer H., Trauzold A., Siegel E.G., Folsch U.R., Schmidt W.E.;				
RT	"PRG1: a novel early-response gene transcriptionally induced by				
RT	pituitary adenylate cyclase activating polypeptide in a pancreatic				
RT	carcinoma cell line.";				
RL	Cancer Res. 56:2641-2641(1996).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=97339426; PubMed=9196025;				
RA	Pietzsch A., Buechler C., Aslanidis C., Schmitz G.;				
RT	"Identification and characterization of a novel monocyte/macrophage				
RT	differentiation-dependent gene that is responsive to				
RT	lipopolysaccharide, ceramide, and lysophosphatidylcholine.";				
RL	Biochem. Biophys. Res. Commun. 235:4-9(1997).				
RL	[4]				
RP	PRELIMINARY SEQUENCE OF 1-106 FROM N.A.				
RC	TISSUE=PLACENTA;				
RA	Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M.,				
RA	Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,				
RA	Rifkin L., Rohlfing T., Tan F., Trevaskis E., Waterston R.,				
RA	Williamson A., Wohlmann P., Wilson R.;				
RL	Submitted (FER-1995) to the EMBL/GenBank/DBJ databases.				
CC	-!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).				
CC	-!- INDUCTION: BY RADIATION, TPA, OKADAIC ACID AND TNF-ALPHA.				
CC	-!- PTM: GLYCOSYLATED.				
CC	-!- SIMILARITY: STRONG, TO MOUSE ORTHOLOG.				
CC	-!- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO				
CC	FRAMESHIFTS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation				

```
DR FLYBASE: FBgn0001965; Sos.
DR INTERPRO: IPR000219; -.
DR INTERPRO: IPR000651; -.
DR INTERPRO: IPR001849; -.
DR INTERPRO: IPR001895; -.
DR PFAM: PF00169; PH; 1.
DR PFAM: PF00617; RasGEF; 1.
DR PFAM: PF00621; RhogEF; 1.
DR PFAM: PF00621; RhogEF; 1.
DR PROSITE: PS00720; GDS_CDC35; 1.
DR PROSITE: PS00003; PH_DOMAIN; 1.
DR KW Guanine-nucleotide releasing factor; Neurogenesis.
FT DOMAIN 15 51
FT DOMAIN 248 478
FT DOMAIN 479 587
FT DOMAIN 1511 1516
FT DOMAIN 1525 1541
FT CONFLICT 232 243
FT CONFLICT 1462 1462
FT SEQUENCE 1595 AA; 177837 MW; 33AE31F0767A219F CRC64;

Query Match 17.88; Score 78.5; DB 1; Length 1595;
Best Local Similarity 28.38; Pred. No. 3.9;
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRRQLPVE-----EPNPAKRLFLLLTIVFCOILMNEGVPAPLPPEDAPNAAASLAPT 53
DB 1472 IRRNSAIKRAAATVSNPNOAAGISTTLTVSOAVATDEVLPPLISP-----AASSSTT 1476

QY 54 --PVSPVLEPF--NLTSPE-----SDXALDLSTFLQO-----HPAAF 86
DB 1477 TSPLTPAMSPNIPSPHPVESTSSSYAHQMRQOQOQOQOQTHPAIY 1522

RESULT 4
GSPC_ERWCH STANDARD; PRT; 272 AA.
AC P31698;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN C (PECTIC ENZYMES SECRETION PROTEIN
DE OUTC).
GN OUTC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC16;
RX MEDLINE=93054355; PubMed=1429461;
RA Lindeberg M., Collmer A.;
RT "Analysis of eight out genes in a cluster required for pectic enzyme
RT secretion by Erwinia chrysanthemi: sequence comparison with secretion
RT genes from other gram-negative bacteria.";
RL J. Bacteriol. 174:7385-7397(1992).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L02214; AAA24830.1; .
```

```
DR PIR: A47021; A47021.
DR INTERPRO: IPR001478; -.
DR INTERPRO: IPR001639; -.
DR PFAM: PF00595; PD2; 1.
DR PRINTS: PRO0810; BCTERIALGSPC.
DR PROSITE: PS01141; T2SP_C; 1.
DR KW Transport; Transmembrane; Inner membrane.
FT DOMAIN 1 16
FT DOMAIN 17 35
FT TRANSMEM 17 35
FT DOMAIN 36 272
FT SEQUENCE 272 AA; 30301 MW; 064C7311F9714405 CRC64;

Query Match 16.68; Score 73.5; DB 1; Length 272;
Best Local Similarity 30.18; Pred. No. 1.7;
Matches 22; Conservative 13; Mismatches 25; Indels 13; Gaps 3;

QY 4 QLPVEPNPAKRLFLLLTIVFCQ---ILMAEAGVPAPLPEDAPNAAASLAPTVPSPVLE 60
DB 5 KLPLSPSVIRILFYLLMLLFCQLAMIFWRVGLP-----DNSPVASVQITPAOAROQ 58
QY 61 PFNLTSEPSDYAL 73
DB 59 PVTL-----NDFTL 67

RESULT 5
GSPC_ERWCH STANDARD; PRT; 272 AA.
ID GSPC_ERWCH
AC Q01564;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN C (PECTIC ENZYMES SECRETION PROTEIN
DE OUTC).
GN OUTC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5937;
RX MEDLINE=93086427; PubMed=1453958;
RA Condemine G., Dorel C., Hugouvieux-Cotte-Pattat N., Robert-Baudouy J.;
RT "Some of the out genes involved in the secretion of pectate lyases in
RT Erwinia chrysanthemi are regulated by kdgr.";
RL Mol. Microbiol. 6:3199-3211(1992).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X65265; CAA46369.1; .
DR PIR: S28013; S28013.
DR INTERPRO: IPR001478; -.
DR INTERPRO: IPR001639; -.
DR PFAM: PF00595; PD2; 1.
DR PRINTS: PRO0810; BCTERIALGSPC.
DR PROSITE: PS01141; T2SP_C; 1.
DR KW Transport; Transmembrane; Inner membrane.
FT DOMAIN 1 16
FT DOMAIN 17 35
FT TRANSMEM 17 35
FT DOMAIN 36 272
FT SEQUENCE 272 AA; 30162 MW; 0E2C1A952B42605A CRC64;
```

CC ACTIN-BINDING PROTEIN.
 CC -1- MISCELLANEOUS: BINDING SITES FOR VON WILLEBRAND FACTOR AND
 CC THROMBIN (THE LATTER SITE WITH UNKNOWN FUNCTION) ARE IN THE
 CC AMINO-TERMINAL PART OF THE MOLECULE.
 CC -1- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
 CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J02940; AA52595.1; -;
 CC EMBL: M22403; AA52596.1; -;
 CC PIR: A27075; NBUHA.
 CC MIM: 231200; -;
 CC MIM: 177820; -;
 CC INTERPRO: IPR000372; -;
 CC INTERPRO: IPR000483; -;
 CC INTERPRO: IPR001611; -;
 CC PFAM: PF00560; LRR: 6;
 CC PFAM: PF01463; LRRCT: 1;
 CC PFAM: PF01462; LRRNT: 1;
 CC PRINTS: PR00019; LEURICHRPT.
 CC Platelet; Transmembrane; Glycoprotein; Hemostasis; Blood coagulation;
 CC Repeat; Leucine-repeat; Signal; Cell adhesion; Disease mutation;
 CC Polymorphism; von Willebrand disease; Bernard Soulier syndrome.
 CC SIGNAL 1 16
 CC CHAIN 17 626 PLATELET GLYCOPROTEIN IB ALPHA CHAIN.
 CC CHAIN 17 626 GLYCOCALICIN.
 CC DOMAIN 17 505 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 506 526 POTENTIAL.
 CC DOMAIN 527 626 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 39 224 LEUCINE-RICH REPEATS.
 CC REPEAT 39 60 LRR 1.
 CC REPEAT 61 83 LRR 2.
 CC REPEAT 84 108 LRR 3.
 CC REPEAT 109 131 LRR 4.
 CC REPEAT 132 155 LRR 5.
 CC REPEAT 156 179 LRR 6.
 CC REPEAT 180 204 LRR 7.
 CC REPEAT 205 224 LRR 8.
 CC REPEAT 379 386 THR/PRO-RICH.
 CC REPEAT 387 395 THR/PRO-RICH.
 CC REPEAT 400 408 THR/PRO-RICH.
 CC REPEAT 409 417 THR/PRO-RICH.
 CC REPEAT 422 430 THR/PRO-RICH.
 CC DISULFID 20 33
 CC DISULFID 225 264
 CC DISULFID 227 280
 CC CARBOHYD 37 37 N-LINKED (GLCNAC. . .).
 CC CARBOHYD 175 175 N-LINKED (GLCNAC. . .).
 CC CARBOHYD 308 308
 CC VARIANT 73 73
 CC VARIANT 161 161
 CC VARIANT 172 172
 CC VARIANT 195 195
 CC VARIANT 225 225
 CC VARIANT 249 249
 CC VARIANT 255 255
 CC VARIANT 626 626
 CC SEQUENCE 626 AA; 68955 MW; 19514119845DF573 CRC64;
 CC
 CC Query Match 16.2%; Score 71.5; DB 1; Length 626;
 CC Best Local Similarity 47.5%; Pred. No. 8.2;
 CC Matches 19; Conservative 2; Mismatches 18; Indels 1; Gaps 1;

Best Local Similarity 45.0%; Pred. No. 6.8;
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;
 QY 36 PAPLP-----PDADNAAASLAPTPVSPVLEPFLNLTSEPS 69
 DB 382 PTPSPPTSEVPPEPAPNMTLTPTSPPTPEP---TSEPA 418
 RESULT 7
 BRLL_EBV
 ID BRLL_EBV STANDARD; PRT: 605 AA.
 AC P03209;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE TRANSCRIPTION ACTIVATOR BRLL1.
 GN BRLL1.
 OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tuffnell P.S., Barrell B.G.;
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RL Nature 310:207-211(1984).
 CC -1- FUNCTION: TRANSCRIPTION ACTIVATION.
 CC -1- MISCELLANEOUS: BRLL1 PROTEIN IS AN EARLY OR IMMEDIATE EARLY
 CC PROTEIN.
 CC -1- SIMILARITY: TO HVS-1 EDRF1 (GENE 50).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: V01555; CAA24814.1; -;
 CC PIR: A03771; Q0BE29.
 CC PIR: S33019; S33019.
 CC TRANSFAC: T00710; -;
 CC Transcription regulation; Activator; DNA-binding; Early protein.
 KW SEQUENCE 605 AA; 56594 MW; B587DF0593FAD808 CRC64;
 SQ
 Query Match 16.0%; Score 70.5; DB 1; Length 605;
 Best Local Similarity 47.5%; Pred. No. 8.2;
 Matches 19; Conservative 2; Mismatches 18; Indels 1; Gaps 1;
 QY 36 PAPLPEDAPNAAASLAPTPVSPVLEPFLNLTSEPSDYALD 74
 DB 453 PGSPWANRPLPASLAPTGTGPVHEPVGSLTPAPVQPLD 492
 RESULT 8
 TRX2_HUMAN
 ID TRX2_HUMAN STANDARD; PRT: 2715 AA.
 AC Q9UMN6; Q9UMN6; Q95836; Q9Y669; Q9Y668; O15022;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE TRITHORAX HOMOLOG 2 (MIXED LINEAGE LEUKEMIA GENE HOMOLOG 2 PROTEIN)
 DE (K1AA0340).
 GN TRX2 OR HRX2 OR MLL2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).


```

RT "Mapping the cleavage site in protein synthesis initiation factor
RT eif-4 gamma of the 2A proteases from human Coxsackievirus and
RL rhinovirus.";
RN J. Biol. Chem. 268:19200-19203(1993).
[2]
RP PARTIAL SEQUENCE..
RX MEDLINE=33034654; Pubmed=1429670;
RY Yan R., Rychlik W., Echolson D., Rhoads R.E.;
RT "Amino acid sequence of the human protein synthesis initiation factor
RT eIF-4 gamma.";
RN J. Biol. Chem. 267:23226-23231(1992).
CC -!- FUNCTION: COMPONENT OF THE PROTEIN COMPLEX EIF-4, WHICH IS
CC INVOLVED IN THE RECOGNITION OF THE MRNA CAP, ATP-DEPENDENT
CC UNWINDING OF 5'-TERMINAL SECONDARY STRUCTURE AND RECRUITMENT OF
CC MRNA TO THE RIBOSOME.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -!- SIMILARITY: THE C-TERMINAL REGION IS SIMILAR TO THE N-TERMINAL
CC REGION OF WHEAT EUKARYOTIC INITIATION FACTOR (ISO)4F SUBUNIT P82.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR ENBL; L22090; AAA31242.1; ..
DR INTERPRO; IPR000504; ..
DR PFAM; PF02020; IF5_eIF4_eif2; 1.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW Initiation factor; Protein biosynthesis; Phosphorylation;
RNA-binding.
FT DOMAIN 188 192 POLY-PRO.
FT DOMAIN 262 275 POLY-GLU.
FT DOMAIN 502 607 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 698 706 RNA-BINDING (RNP1) (BY SIMILARITY).
FT DOMAIN 1393 1398 POLY-GLU.
SQ SEQUENCE 1402 AA; 154050 MW; 7FD85D7E30519230 CRC64;
-----
Query Match 15.6%; Score 69; DB 1; Length 1402;
Best Local Similarity 30.2%; Pred.No. 29;
Matches 26; Conservative 7; Mismatches 21; Indels 32; Gaps 4;

Oy 4 QLPVEERNPAKRLLFLTLITVFCOILMAECVPAPLPEDPAPNAAASLAPTVVS-PVLEPF 62
   |::||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 100 QMPVEEPAPISR-----EAGEPYCLSP-----PTPLAEPILEVE 134
                                     |||:: |::|

Oy 63 NLTSEP-----SDYALDLSTFLQOH 82
   |:: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 VTLKPVDPVSEFFSSPQVLTPASH 160
                               STANDARD;          PTG;    181 AA.

RESULT 11
C21O_HUMAN
ID C21O_HUMAN STANDARD;          PTG;    181 AA.
AC Q9Y426;
DT 30-MAY-2000 (Rel. 39, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROTEIN C21ORF25 (FRAGMENT).
CN C21ORF25;
CS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Koehrer K., Beyer A., Meves H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (May-1999) to the EMBL/GenBank/DDNJ databases.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```



```

CC -----
EMBL: D12686; BAA02185.1; -.
DR DR
MIM: 600495; -.
DR DR
INTERPRO: IPR000504; -.
DR DR
PFAM: PF02020; IF5_eif4_eif2; 1.
DR DR
PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
DR DR
KW Initiation factor; Protein biosynthesis; Phosphorylation;
DR RNA-binding.
FT FT
FT DOMAIN 258 271 POLY-GLU.
FT FT
FT DOMAIN 305 308 POLY-ALA.
FT FT
FT DOMAIN 599 605 RNA-BINDING (RNP2) (BY SIMILARITY).
FT FT
FT DOMAIN 696 704 RNA-BINDING (RNP1) (BY SIMILARITY).
FT FT
FT CONFLICT 497 500 PRGP -> ARGAO (IN REF. 1).
SQ SQ
SEQUENCE 1395 AA; 153360 MW; C712A09D270DF2A8 CRC64;

```

Search completed: January 30, 2001, 23:12:25
Job time: 4006 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 12:11:54 ; Search time 1900.13 Seconds
(without alignments)
3307.456 Million cell updates/sec

Title: us-08-799-910-9

Perfect score: 1228

Sequence: 1 ATGTGCTACTCTCGACGCTG.....AAAAAAAAAAAACTCGAG 1228

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_bal:*
- 2: gb_baz:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_ph:*
- 6: gb_pl1:*
- 7: gb_pl2:*
- 8: gb_pl3:*
- 9: gb_pr2:*
- 10: gb_pr3:*
- 11: gb_ro:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: em_fun:*
- 15: em_hum1:*
- 16: em_hum2:*
- 17: em_in:*
- 18: em_om:*
- 19: em_or:*
- 20: em_ov:*
- 21: em_pat:*
- 22: em_ph:*
- 23: em_pi:*
- 24: em_ro:*
- 25: em_sts:*
- 26: em_sy:*
- 27: em_un:*
- 28: em_vi:*
- 29: gb_ba3:*
- 30: gb_in1:*
- 31: gb_in2:*
- 32: gb_in3:*
- 33: gb_pi3:*
- 34: gb_pr4:*
- 35: em_bal:*
- 36: em_ba2:*
- 37: em_htg1:*
- 38: em_htg2:*
- 39: em_htg3:*
- 40: em_htg4:*
- 41: em_htg5:*
- 42: em_htg6:*
- 43: em_htg7:*

44: em_htg8:*

45: em_htg9:*

46: em_htg10:*

47: em_hum3:*

48: em_hum4:*

49: em_hum5:*

50: em_hum6:*

51: gb_pr5:*

52: gb_pr6:*

53: gb_pr7:*

54: gb_htg1:*

55: gb_htg2:*

56: gb_htg3:*

57: gb_htg4:*

58: gb_htg5:*

59: gb_htg6:*

60: gb_htg7:*

61: gb_htg8:*

62: gb_htg9:*

63: gb_htg10:*

64: gb_htg11:*

65: gb_htg12:*

66: gb_htg13:*

67: gb_htg14:*

68: gb_htg15:*

69: gb_htg16:*

70: gb_htg17:*

71: gb_htg18:*

72: gb_htg19:*

73: gb_htg20:*

74: gb_htg21:*

75: gb_htg22:*

76: gb_htg23:*

77: gb_sts1:*

78: gb_sts2:*

79: gb_v11:*

80: gb_v12:*

81: gb_pat1:*

82: gb_pat2:*

83: em_htg0:*

84: gb_htg24:*

85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match %	Length	DB ID	Description
1	1207.8	98.4	1230	53	HSD1P2
2	1152.2	93.8	1223	85	S81914
3	1080.4	88.0	44118	9	AC006165
4	1080.4	88.0	90244	8	AB023051
5	1080.4	88.0	200000	51	AP000512
6	1048.6	85.4	1309	10	AF039067
7	1048.6	85.4	1693	10	AF071596
8	979.8	79.8	1864	53	HSPRGI
9	471	38.4	477	52	AF083421
10	298.8	24.3	343	52	HSA227914
11	242.4	19.7	1758	11	RNPRG1
12	210.6	17.1	203476	70	AC074150
13	209	17.0	1938	11	MMGLY96
14	202	16.4	187868	62	AC022301
15	151	12.3	297	81	A75448
16	151	12.3	297	81	A78427
17	122	9.9	122	78	G82116
18	74	6.0	205512	74	AL355388
19	70.4	5.7	267	81	A75468
20	70.4	5.7	267	81	A78447
21	64	5.2	83276	65	AC026631

Db	372	GGCCACCCCTGTGTCCCTCTCGAGCCCTTAATCTGACTTCGGAGCCCTCGGAC	431
Qy	421	TAGCTCTGGACCTGAGCACTTCTCTCCAGCAACACCCCGCCCTTCTAACTGTGACTC	480
Db	432	TAGCTCTGGACCTGAGCACTTCTCTCCAGCAACACCCCGCCCTTCTAACTGTGACTC	491
Qy	481	CCGCGACTCCCCAANAAGATCCGAAAAACCAAAAGAACACACAGCGGTACTGTGGG	540
Db	492	CCGCGACTCCCCAANAAGATCCGAAAAACCAAAAGAACACACAGCGGTACTGTGGG	551
Qy	541	CGAGAGCGTATCCCAACTGGGACTTCCAGGCAACTTCAACTCAGAACACTACAGCGA	600
Db	552	CGAGAGCGTATCCCAACTGGGACTTCCAGGCAACTTCAACTCAGAACACTACAGCGA	611
Qy	601	GAGCCACCCCGTCTGTAGGCGGAGCCAGGCGCACAGAGCCAGCGCATAGAGACC	660
Db	612	GAGCCACCCCGTCTGTAGGCGGAGCCAGGCGCACAGAGCCAGCGCATAGAGACC	671
Qy	661	GAGGCACAGCCAGCTGGGGCTAGGCCGCTGGGAAGGAGAGCGTCTTAATTTATTTCT	720
Db	672	GAGGCACAGCCAGCTGGGGCTAGGCCGCTGGGAAGGAGAGCGTCTTAATTTATTTCT	731
Qy	721	TATTGCTCTTAATTAATATTATCTATTATGTAAGTCTCTCTAGGTGATGGAGATCT	780
Db	732	TATTGCTCTTAATTAATATTATCTATTATGTAAGTCTCTCTAGGTGATGGAGATCT	791
Qy	781	GTACGTAATATTATTTAACTTATCAAGGGTGTGAGATGTTCCCTCTGCTGTAATGC	840
Db	792	GTACGTAATATTATTTAACTTATCAAGGGTGTGAGATGTTCCCTCTGCTGTAATGC	851
Qy	841	AGTCTCTTGGTATTTATGAGCTTTGTGGGACTGTGGAAGCAGACACTGGAAGTGC	900
Db	852	AGTCTCTTGGTATTTATGAGCTTTGTGGGACTGTGGAAGCAGACACTGGAAGTGC	911
Qy	901	GGCAAGTACGACAAGAAATGGGAGACTTCGGGTGGGGAGGAGCTCCGGCTGGGATG	960
Db	912	GGCAAGTACGACAAGAAATGGGAGACTTCGGGTGGGGAGGAGCTCCGGCTGGGATG	971
Qy	961	AAGTCTGCTGGTGGGCTGAAGTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCG	1020
Db	972	AAGTCTGCTGGTGGGCTGAAGTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCG	1031
Qy	1021	TCGTCTACTGTGAGACTTCGGCGGACCATTAGGATGAGATCGGTGAGATCTCTCCA	1080
Db	1032	TCGTCTACTGTGAGACTTCGGCGGACCATTAGGATGAGATCGGTGAGATCTCTCCA	1091
Qy	1081	TCCTTTGAAGTCCCTTTAGGTTAGGCTGCGAGTACAGGTTGGGGTTGGTGGGCTGT	1140
Db	1092	TCCTTTGAAGTCCCTTTAGGTTAGGCTGCGAGTACAGGTTGGGGTTGGTGGGCTGT	1151
Qy	1141	CACGAGCGACTGTGAGATCGGCTAGTATGTTCTGTGAACACAAATAAATTCATTAC	1200
Db	1152	CACGAGCGACTGTGAGATCGGCTAGTATGTTCTGTGAACACAAATAAATTCATTAC	1211
Qy	1201	TGTCAAAAAATAAATAA 1219	
Db	1212	TGTCGTAAAAAATAAATA 1230	
RESULT 2			
LOCUS	S81914	1223 bp	mRNA
DEFINITION	LEX-1-radiation-inducible immediate-early gene [human, placenta, mRNA partial, 1223 nt].		
ACCESSION	S81914		
VERSION	S81914.1	GI:1488384	
KEYWORDS	human placenta.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1223)		
AUTHORS	Kondratyev, A.D., Chung, K.N. and Jung, M.O.		

TITLE	Identification and characterization of a radiation-inducible glycosylated human early-response gene		
JOURNAL	Cancer Res. 56 (7), 1498-1502 (1996)		
MPDLINE	96181295		
REMARK	GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 177109] from the original journal article. This sequence comes from Fig. 2A.		
FEATURES	Location/Qualifiers		
source	1..1223		
gene	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
CDS	19..489		
	/gene="LEX-1"		
	/note="radiation-inducible immediate-early gene"		
	19..489		
	/gene="LEX-1"		
	/note="radiation-inducible immediate-early gene; This sequence comes from Fig. 2A"		
	/codon_start=1		
	/protein_id="AAB36278.1"		
	/db_xref="GI:1488385"		
	/translation="MCHSRSCHPMTILOATPPAPSTIPGPRRSGPEITFDNLPPEP		
	AAAPAGRPSGSRGHRKRVLYPRVRRPLPVRPNPAKRLFLLLTVFCOLLMAE		
	EGVRAPLPEDAPNAASLAPTPVSPVLEPNLTSEPSYALDLSTFLOQHPANP"		
BASE COUNT	249 a	368 c	342 g
ORIGIN	264 t		
Query Match	93.8%	Score 1152.2	DB 85; Length 1223;
Best Local Similarity	99.4%	Pred. No. 3.7e-255;	
Matches 1198;	Conservative	0; Mismatches	3; Indels 4; Gaps 4;
Qy	1	ATGTGTCACTCTCGAGCTGCCACCGACCATGACCATCTCGAGGCCCGGACCCGGCC	60
Db	19	ATGTGTCACTCTCGAGCTGCCACCGACCATGACCATCTCGAGGCCCGGACCCGGCC	78
Qy	61	CTCTCCACCATCCCGGAGCCCGCGGGCTCGGTCTCTGAGATCTTCACCTTCGACCT	120
Db	79	CTCTCCACCATCCCGGAGCCCGCGGGCTCGGTCTCTGAGATCTTCACCTTCGACCT	138
Qy	121	CTCCCGGAGCCGACGCGCCCTCGCGGGCCCGGAGCCCTCTCGCGGGCACCCAAAG	180
Db	139	CTCCCGGAGCCGACGCGCCCTCGCGGGCCCGGAGCCCTCTCGCGGGCACCCAAAG	198
Qy	181	CGCAGCCGAGGCTTCTTACCTTCAGTTCGTCGCGGCGCAGTTCGCGAGTTCGAGAACCG	240
Db	199	CGCAGCCGAGGCTTCTTACCTTCAGTTCGTCGCGGCGCAGTTCGCGAGTTCGAGAACCG	258
Qy	241	AACCCAGCCAAAGGCTTCTTCTTCTGCTCTCACCATCTCTCTCTGCGCAGATCTTGATG	300
Db	259	AACCCAGCCAAAGGCTTCTTCTTCTGCTCTCACCATCTCTCTCTGCGCAGATCTTGATG	318
Qy	301	GCTGAAGAGGCTGTCGCGGCGCCCTCGCTCCAGAGGAGCCCTAACCGCCGATCCCTG	360
Db	319	GCTGAAGAGGCTGTCGCGGCGCCCTCGCTCCAGAGGAGCCCTAACCGCCGATCCCTG	378
Qy	361	GCGCCACCCCTGTGTCTCCCGCTCTCGAGCCCTTTAATCTGACTTCGAGCCCTTCGGAC	420
Db	379	GCGCCACCCCTGTGTCTCCCGCTCTCGAGCCCTTTAATCTGACTTCGAGCCCTTCGGAC	438
Qy	421	TAGCTCTGGACCTGAGCACTTCTCTCCAGCAACACCCCGCCCTTCTAACTGTGACTC	480
Db	439	TAGCTCTGGACCTGAGCACTTCTCTCCAGCAACACCCCGCCCTTCTAACTGTGACTC	498
Qy	481	CCGCGACTCCCCAANAAGATCCGAAAAACCAAAAGAACACACAGCGGTACTGTGGG	540
Db	499	CCGCGACTCCCCAANAAGATCCGAAAAACCAAAAGAACACACAGCGGTACTGTGGG	558
Qy	541	CGAGAGCGTATCCCAACTGGGACTTCCGAGGCAACTTGAACCTCAGAACACTACACCGGA	600
Db	559	CGAGAGCGTATCCCAACTGGGACTTCCGAGGCAACTTGAACCTCAGAACACTACACCGGA	618
Qy	601	GAGCCACCCGCTGTGTAGGCGGAGCCGAGGCGCACAGAGCCGAGGCGCATAGAGACC	660

```
repeat_region 3941..4070
repeat_region /rpt_family="Alu"
repeat_region complement(4449..4724)
repeat_region /rpt_family="Alu"
repeat_region complement(5117..5389)
repeat_region /rpt_family="Alu"
repeat_region 7668..7913
repeat_region /rpt_family="Alu"
repeat_region complement(7924..8226)
repeat_region /rpt_family="Alu"
repeat_region 8250..8551
repeat_region /rpt_family="Alu"
repeat_region complement(8893..8973)
repeat_region /rpt_family="MIR"
repeat_region complement(11829..12096)
repeat_region /rpt_family="Alu"
repeat_region 12416..12740
repeat_region /rpt_family="Alu"
repeat_region complement(13181..13247)
repeat_region /rpt_family="Alu"
repeat_region complement(13498..13647)
repeat_region /rpt_family="L1"
repeat_region 13869..14279
repeat_region /rpt_family="Alu"
repeat_region 18033..18284
repeat_region /rpt_family="Alu"
repeat_region 19178..19451
repeat_region /rpt_family="Alu"
repeat_region 19458..19749
repeat_region /rpt_family="Alu"
repeat_region 20862..21002
repeat_region /rpt_family="MER5"
repeat_region 21566..21832
repeat_region /rpt_family="Alu"
repeat_region 22231..22521
repeat_region /rpt_family="Alu"
repeat_region complement(22616..22679)
repeat_region /rpt_family="Alu"
repeat_region complement(22869..23145)
repeat_region /rpt_family="Alu"
repeat_region complement(23478..23772)
repeat_region /rpt_family="Alu"
repeat_region 24184..24372
repeat_region /rpt_family="Alu"
repeat_region 24529..24813
repeat_region /rpt_family="Alu"
repeat_region complement(25060..25342)
repeat_region /rpt_family="Alu"
repeat_region complement(25367..25653)
repeat_region /rpt_family="Alu"
repeat_region 25787..26076
repeat_region /rpt_family="Alu"
repeat_region 26491..26762
repeat_region /rpt_family="Alu"
repeat_region 26866..27143
repeat_region /rpt_family="Alu"
repeat_region complement(27620..27919)
repeat_region /rpt_family="Alu"
repeat_region 29525..29719
repeat_region /rpt_family="MIR"
repeat_region 29762..30333
repeat_region /rpt_family="Alu"
repeat_region 30365..30653
repeat_region /rpt_family="Alu"
repeat_region 31430..31722
repeat_region /rpt_family="Alu"
repeat_region 31765..31857
repeat_region /rpt_family="Alu"
repeat_region 32265..32563
repeat_region /rpt_family="Alu"
repeat_region complement(32710..32858)
repeat_region /rpt_family="Alu"
repeat_region complement(33374..33584)
```

```
repeat_region /rpt_family="MER2"
repeat_region complement(34547..34595)
repeat_region /rpt_family="MIR"
repeat_region complement(36673..36954)
repeat_region /rpt_family="Alu"
repeat_region complement(37002..37230)
repeat_region /rpt_family="Alu"
repeat_region complement(40018..40170)
repeat_region /rpt_family="Alu"
variation 40452
note="high quality variation versus 3' overlapping clone"
repeat_region 40509..41109
repeat_region /rpt_family="Alu"
repeat_region 42609..42892
repeat_region /rpt_family="Alu"
repeat_region 43240..43370
repeat_region /rpt_family="MIR"
repeat_region 43493..43779
repeat_region /rpt_family="Alu"
BASE COUNT 11505 a 10560 c 11226 g 10827 t
ORIGIN
Query Match 88.0%; Score 1080.4; DB 9; Length 44118;
Best Local Similarity 91.4%; Pred. No. 1e-238;
Matches 1203; Conservative 0; Mismatches 1; Indels 112; Gaps 1;
QY 1 ATGTGTCACCTCTGCAGCTGCCAGCCCGACCATGACCATCTCTGAGAGCCCGACCCGCGCC 60
DB 14601 ATGTGTCACCTCTGCAGCTGCCAGCCCGACCATGACCATCTCTGAGAGCCCGACCCGCGCC 14660
QY 61 CCCTCCACCATCCCGGGGACCCCGGGGCTCCGGTCTGAGATCTTCACCTTCGACCT 120
DB 14661 CCCTCCACCATCCCGGGGACCCCGGGGCTCCGGTCTGAGATCTTCACCTTCGACCT 14720
QY 121 CTCGCGAGCCCGCAGCGGCCCTTCCGCGGCCCGCCGCGCTCTCGCGGCGCACCGAAG 180
DB 14721 CTCGCGAGCCCGCAGCGGCCCTTCCGCGGCCCGCCGCGCTCTCGCGGCGCACCGAAG 14780
QY 181 CGAGCGCGAGGGTTCTTACCTCGAGT----- 209
DB 14781 CGAGCGCGAGGGTTCTTACCTCGAGTGTGAGTATCGCGCGAAGTGGGCAITTCGCGGG 14840
QY 210 ----- 209
DB 14841 GTGGGTGCCCTGGAGTCACTGGGGAGACCGCCGACTCCAGAGGCTCTGACCTGACCTGT 14900
QY 210 -----GGTCCGCGCGCCAGCTGCCACTCGAGGAACCGAACCCAGCC 248
DB 14901 CTCCTGTTTGTCTCTCCCTTAGGTCCGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCC 14960
QY 249 CAAAGGCTTCTTTTCTGTGCTCAGCATCGTCTTTCGCGAGATCTCTGATGGTGAAGA 308
DB 14961 CAAAGGCTTCTTTTCTGTGCTCAGCATCGTCTTTCGCGAGATCTCTGATGGTGAAGA 15020
QY 309 GGGTGTCCGCGCGCCCTGCTCCAGAGGAGCCCGCTTAACCGCGCATCCCTCGGCGCCAC 368
DB 15021 GGGTGTCCGCGCGCCCTGCTCCAGAGGAGCCCGCTTAACCGCGCATCCCTCGGCGCCAC 15080
QY 369 CCCTGTGTCCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGGCCCTCGGACTACGCTCT 428
DB 15081 CCCTGTGTCCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGGCCCTCGGACTACGCTCT 15140
QY 429 GGACCTCAGCACTTTCTCTCCAGCAACACCGCGCGCTTTAATCTGACTTCGCGGCACT 488
DB 15141 GGACCTCAGCACTTTCTCTCCAGCAACACCGCGCGCTTTAATCTGACTTCGCGGCACT 15200
QY 489 CCCCAAAAGAAATCCGAAACACCAAGAAACACCGAGGCTTCTGTCGCGGAGAGCG 548
DB 15201 CCCCAAAAGAAATCCGAAACACCAAGAAACACCGAGGCTTCTGTCGCGGAGAGCG 15260
QY 549 TATCCCCCAACTGGGACTTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGAGCGCCAC 608
DB 549 TATCCCCCAACTGGGACTTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGAGCGCCAC 608
```

```
Db 61518 CTATTAATATTTATATGTTATTTATGATCTCTCTAGTGTGAGATGTGACGTAA 61577
Qy 789 TATTATTATTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTGTAATGCAAGTCTCT 848
Db 61578 TATTATTATTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTGTAATGCAAGTCTCT 61637
Qy 849 TGTATTATTAGCTTTGTGGGACTTGTGGAGCAGGACACTGCAACTGCGGCAAACT 908
Db 61638 TGTATTATTAGCTTTGTGGGACTTGTGGAGCAGGACACTGCAACTGCGGCAAACT 61697
Qy 909 AGGAGAAGAAATCGGGAGGACTCGGCTGGGGAGGAGACGTCCCGGCTGGGATGAAGTCTCG 968
Db 61698 AGGAGAAGAAATCGGGAGGACTCGGCTGGGGAGGAGACGTCCCGGCTGGGATGAAGTCTCG 61757
Qy 969 TGTGGGCTGTAAGTTTAGGAGGTGACTGCACTCCAGCAGATCTCAACTCCGCTCTCTA 1028
Db 61758 TGTGGGCTGTAAGTTTAGGAGGTGACTGCACTCCAGCAGATCTCAACTCCGCTCTCTA 61817
Qy 1029 CTGTGTGAGACTTCGGCGGACCAATTAGGAATGAGATCCGTGAGATCCTTCCATCTCTTG 1088
Db 61818 CTGTGTGAGACTTCGGCGGACCAATTAGGAATGAGATCCGTGAGATCCTTCCATCTCTTG 61877
Qy 1089 AACTCGCCTTTAGGTGGCTCGGAGGTAGAGGTTGGGGTTGGTGGCTGTGACGAGC 1148
Db 61878 AACTCGCCTTTAGGTGGCTCGGAGGTAGAGGTTGGGGTTGGTGGCTGTGACGAGC 61937
Qy 1149 GACTGTGAGATCGCCTAGTATGTTCTGTGAACACAAATAAATGATTTACTGTC 1204
Db 61938 GACTGTGAGATCGCCTAGTATGTTCTGTGAACACAAATAAATGATTTACTGTC 61993

RESULT 5
LOCUS AP000512
DEFINITION Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
section 11/20.
ACCESSION AP000512
VERSION AP000512.1 GI:5926699
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Shinya.S., Tamiya.G., Oka.A. and Inoko.H.
TITLE Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region
JOURNAL Published Only in DataBase (1999) In press
REFERENCE
AUTHORS Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
TITLE Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Miki
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp.
URL:http://www.alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is conducted by Tokai University as a JST sequencing
Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884.
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive web site (http://www.alis.tokyo.jst.go.jp/HGS/Lop.
html) or send email to webmaster@www.alis.tokyo.jst.go.jp.
FEATURES
source
1. .200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
```

```
STS 101440. .101648
/note="s1SG4284:The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:4566033"
complement(114197..114328)
/standard_name="D6S2115"
/note="SHGC-13451:The location is between each flanking
site of PCR primers."
STS /db_xref="GDB:738998"
complement(114300..114430)
/note="WI-18721:The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:4574312"
117236..117362
/note="SGC34396:The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:4580073"
complement(117712..117836)
/standard_name="D6S2041"
/note="SHGC-10364:The location is between each flanking
site of PCR primers."
STS /db_xref="GDB:735115"
complement(117961..118093)
/note="SGC32809:The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:4584619"
142184..142339
/note="K1AA0170:The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:4586652"
complement(142215..142343)
/note="WI-15445:The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:4581086"
complement(153977..154114)
/note="SGC34551:The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:4582352"
complement(165191..165295)
/standard_name="D6S1376"
/note="downstream/upstream:The location is between each
flanking site of PCR primers."
STS /db_xref="GDB:580497"
188925..188989
/note="pml216:The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:594325"
complement(197650..197804)
/note="cda01e04:The location is between each flanking site
of PCR primers."
BASE COUNT 52458 a 48336 c 49153 g 50053 t
ORIGIN
Query Match 88.0%; Score 1080.4; DB SI; Lenqth 200000;
Best Local Similarity 91.4%; Pred. No. 9.6e-239;
Matches 1203; Conservative 1; Mismatches 112; Gaps 1;
Qy 1 ATGTGTCACTCTCGGAGCTGCCACCGGACCATGACCATCTCGAGCCCGCCGGCC 60
Db 97663 ATGTGTCACTCTCGGAGCTGCCACCGGACCATGACCATCTCGAGCCCGCCGGCC 97722
Qy 61 CCTCCACCATCCCGGAGCCCGGGGGCTCCGGTCTCTGAGATCTTCACTTTCGACCT 120
Db 97723 CCTCCACCATCCCGGAGCCCGGGGGCTCCGGTCTCTGAGATCTTCACTTTCGACCT 97782
Qy 121 CTCCTGGAGCCCGAGCGGCCCTTCGCCGGCGCCCGCCAGCGCCCTCTCGGGGACCAAG 180
Db 97783 CTCCTGGAGCCCGAGCGGCCCTTCGCCGGCGCCCGCCAGCGCCCTCTCGGGGACCAAG 97842
Qy 181 CGCAGCGGAGGTTCTCTACCTCGAGT----- 209
```


QY 312 TGTGCGGCGGCGCTGCTCCAGAGAGAGCGCCCTACGCGCGATCC-----T 359
Db 1132 TGTGTCGACGCGCTGGCTCCGAGGATGCTACGAGCGCGGTGACACCTGAGCCCATTC 1191
QY 360 GCGCGCCACCGCTGTGCTCCCGCTCTCTCAGCGCTTAACTGACTTCGGAGCCCTCGGA 419
Db 1192 TGGGCCATTACTGCGCGCGCGCTCTCGAGCTTTGACCTGACTCGGAGTCTCGGA 1251
QY 420 CTACGCTTGGACCTGACGACTTCTCCAGCAACACCGCGCGCTTCTAACTGTGACT 479
Db 1252 CTATGCGCTGGATCTTAAAGCTTTCTTAGCAACATCCGCGCGCTTCTAAACGCGATG 1311
QY 480 CCGCGACTCCCAAAAGATCCGAAACACCAACCAAGAACACAGCGCTACCTGGTCC 539
Db 1312 GGTG-----ACAGTCCGAGAAACAAAGGACCATGGATGGTGGTGGTCC 1358
QY 540 GCGAGAGCTATCCCACTGGGACTTCCGAGCGACTTGAACCTCAGAACACTACAGCGG 599
Db 1359 GAGAGACGCTATCCCAACTGGATTCTAAGCAACGCTAACTCAGAACACTACCGCA 1418
QY 600 AGACGCCACCGGTGTTG 618
Db 1419 AGAGACCGCGGGTCTG 1437

RESULT 14
AC022301
LOCUS 187868 bp DNA HTG 15-JUN-2000
DEFINITION Mus musculus clone RP23-329K24, *** SEQUENCING IN PROGRESS ***, 59
unordered pieces.
ACCESSION AC022301
VERSION AC022301.9 GI:8493517
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 187868)
Mctzker,M.L., Lewis,S., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 187868)
Worley,K.C.
Direct Submission
Submitted (29-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 13, 2000 this sequence version replaced gi:8248590.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: MAA
Center clone name: RP23-329K24
----- Summary Statistics

Sequencing vector: M13; L08021

Chemistry: Dye-primer Bodipy: 82% of reads
Chemistry: Dye-terminator Big Dye: 18% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 13377 bases at least Q40
Consensus quality: 159061 bases at least Q30
Consensus quality: 169688 bases at least Q20
Estimated insert size: 166938; sum-of-contrigs estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contrigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 7154: contrig of 7154 bp in length
7155 7254: gap of unknown length
7255 14340: contrig of 7086 bp in length
14341 14440: gap of unknown length
14441 22916: contrig of 8476 bp in length
22917 23016: gap of unknown length
23017 30834: contrig of 7818 bp in length
30835 30935: gap of unknown length
30936 37993: contrig of 7059 bp in length
37994 38093: gap of unknown length
38094 45511: contrig of 7418 bp in length
45512 45611: gap of unknown length
45612 51842: contrig of 6231 bp in length
51843 51943: gap of unknown length
51944 58572: contrig of 6630 bp in length
58573 58672: gap of unknown length
58673 62718: contrig of 4046 bp in length
62719 62818: gap of unknown length
62819 67974: contrig of 5136 bp in length
67975 68074: gap of unknown length
68075 72883: contrig of 4809 bp in length
72884 72984: gap of unknown length
72985 76726: contrig of 3743 bp in length
76727 76826: gap of unknown length
76827 79331: contrig of 2805 bp in length
79332 79731: gap of unknown length
79732 82858: contrig of 3127 bp in length
82859 82958: gap of unknown length
82959 86054: contrig of 3096 bp in length
86055 86154: gap of unknown length
86155 90283: contrig of 4129 bp in length
90284 90383: gap of unknown length
90384 94598: contrig of 4215 bp in length
94599 99014: contrig of 4316 bp in length
99015 99114: gap of unknown length
99115 103295: contrig of 4181 bp in length
103296 103396: gap of unknown length
103397 106698: contrig of 3303 bp in length
106699 106798: gap of unknown length
106799 106799: contrig of 2859 bp in length
106799 109757: gap of unknown length
109758 113352: contrig of 3635 bp in length
113353 113393: gap of unknown length
113394 117011: contrig of 3519 bp in length
117012 120036: contrig of 2925 bp in length
120037 120136: gap of unknown length
120137 122449: contrig of 2313 bp in length
122450 122549: gap of unknown length
122550 124756: contrig of 1726 bp in length
124757 124766: gap of unknown length
124767 126297: contrig of 1921 bp in length
126298 126397: gap of unknown length
126398 129280: contrig of 2784 bp in length
129281 129280: gap of unknown length

BASE COUNT 66 a 117 c 61 g 52 t 1 others
ORIGIN /db_xref="taxon:32644"

Query Match 12.3%; Score 151; DB 81; Length 297;
Best Local Similarity 76.0%; Pred. No. 1.6e-24;
Matches 212; Conservative 0; Mismatches 65; Indels 2; Gaps 2;
Qy 1 ATGTGTCACTCTGCGACGTGCGACCGACCATGACCATCTGCGAGGCGCCGACCGCGCC 60
Db 16 ATGGAACACTCTGCGGCTGCCACTCCCGCATGACCACTTACAGGCCCTAACCCGAGCC 75
Qy 61 CCTTCCACCATGCCGCGACCGCGGGGCTCGGCTCTGAGATCTTACCTTCGACCCCT 120
Db 76 CCTCAGCTTATCGTCCAGCCACAATA-GCTGCGGCGCTGAGATCTCCACCTCAGACCCCT 134
Qy 121 CTCCCGAGCGCGAGCGGCCCTGCGCGGCGGCCCGCCAGCCCTCTCTCGCGGCGACCGAAAG 180
Db 135 CTCCCGAAGCCACAGCGGCCCTGCGCGGCGACCCCAAGTTC-CTCGCAGGCGACCAAAAG 193
Qy 181 CGCAGCGCGAGGTTCTTACCTCGAGTGTGCGGCGCGAGCTGCCAGTCCGAGGNAACCG 240
Db 194 CGCAGTTCAGGAATCTTACCTTAGAGTATCGGCGCGAGTTGCCAGTTAAGGGACCA 253
Qy 241 AACCCAGCCAAAGGCTTCTTCTTCTGCTGCTCACCATC 279
Db 254 AACCTCGCCAGAGGCTTCTTTTCGGGTGATCGCCATC 292

Search completed: January 30, 2001, 19:23:39
Job time: 25905 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 19:23:39 ; Search time 1900.13 Seconds
(without alignments)
694.889 Million cell updates/sec

Title: US-08-799-910-9_COPY_211_468

Perfect score: 258

Sequence: 1 GTCCGGCCAGCTCCAGT.....AGCAACACCGCGCGCCTTC 258

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 255875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba1.*

2: gb_ba2.*

3: gb_om.*

4: gb_ov.*

5: gb_ph.*

6: gb_pl1.*

7: gb_pl2.*

8: gb_pri.*

9: gb_pr2.*

10: gb_pr3.*

11: gb_ro.*

12: gb_sy.*

13: gb_un.*

14: em_fun.*

15: em_hum1.*

16: em_hum2.*

17: em_in.*

18: em_om.*

19: em_or.*

20: em_ov.*

21: em_pat.*

22: em_ph.*

23: em_pl.*

24: em_ro.*

25: em_sts.*

26: em_sy.*

27: em_un.*

28: em_v1.*

29: gb_ba3.*

30: gb_in1.*

31: gb_in2.*

32: gb_in3.*

33: gb_pl3.*

34: gb_pr4.*

35: em_ba1.*

36: em_ba2.*

37: em_htg1.*

38: em_htg2.*

39: em_htg3.*

40: em_htg4.*

41: em_htg5.*

42: em_htg6.*

43: em_htg7.*

44: em_htg8.*
45: em_htg9.*
46: em_htg10.*
47: em_hum3.*
48: em_hum4.*
49: em_hum5.*
50: em_hum6.*
51: gb_pr5.*
52: gb_pr6.*
53: gb_pr7.*
54: gb_htg1.*
55: gb_htg2.*
56: gb_htg3.*
57: gb_htg4.*
58: gb_htg5.*
59: gb_htg6.*
60: gb_htg7.*
61: gb_htg8.*
62: gb_htg9.*
63: gb_htg10.*
64: gb_htg11.*
65: gb_htg12.*
66: gb_htg13.*
67: gb_htg14.*
68: gb_htg15.*
69: gb_htg16.*
70: gb_htg17.*
71: gb_htg18.*
72: gb_htg19.*
73: gb_htg20.*
74: gb_htg21.*
75: gb_htg22.*
76: gb_htg23.*
77: gb_sts1.*
78: gb_sts2.*
79: gb_v11.*
80: gb_v12.*
81: gb_pat1.*
82: gb_pat2.*
83: em_htg0.*
84: gb_htg24.*
85: gb_pr8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	258	100.0	477	10	AF083421	AF083421 Homo sapi
2	258	100.0	1230	53	HSDF2	Y14551 Homo sapien
3	258	100.0	1309	10	AF039067	AF039067 Homo sapi
4	258	100.0	1693	10	AF071596	AF071596 Homo sapi
5	258	100.0	1864	53	HSPRG1	X96438 H.sapiens P
6	258	100.0	44118	9	AC006165	AC006165 Homo sapi
7	258	100.0	90244	8	AB023051	AB023051 Homo sapi
8	258	100.0	200000	51	AP000512	AP000512 Homo sapi
9	256.4	99.4	1223	85	S81914	S81914 IEX-1-radia
10	154.4	59.8	1758	11	RNPRG1	X96437 R.norvegicu
11	151.2	58.6	203476	70	AC074150	AC074150 Mus muscu
12	149.6	58.0	1938	11	MGGLY96	X67644 M.musculus
13	139.2	54.0	187868	62	AC022301	AC022301 Mus muscu
14	44	17.1	297	81	A75448	A75448 Sequence 11
15	44	17.1	297	81	A78427	A78427 Sequence 11
16	43.2	16.7	115468	57	AC011820	AC011820 Homo sapi
17	43	16.7	7218	81	I66494	I66494 Sequence 14
18	40.2	15.6	34725	29	SC1B2	AL356812 Streptomy
19	39.8	15.4	3897	2	MXU40656	U40656 Myxococcus
20	39.6	15.3	33818	2	MTCY159	283863 Mycobacteri
21	39.4	15.3	2003	85	HUMDAFC1	M64356 Human decay

```

/db_xref="taxon:9606"
/cell_type="leukocytes"
<1..11
/gene="dif-2"
1..1230
/gene="dif-2"
12..482
/gene="dif-2"
/codon_start=1
/product="DIF-2 protein"
/protein_id="CAA74886.1"
/db_xref="GI:2660551"
/db_xref="SWISS-PROT:P46695"
/translation="MCHSRSCPTMTILQAPTAPSTIPGPRRSGPPTFTDPLPEP
AAPAGRPSASGRHKRRSRVLYPRVRRQLPVEENPAKRLLELLLTIVFCQILMAE
EGVAPALPPEDAPNAASLAPTVPVSPVLEPNTSEPSYALDLSFLQOHPAF"
172
/gene="dif-2"
/citation={3}
/replace="g"
328
/gene="dif-2"
/citation={3}
/replace="g"
483..>1230
/gene="dif-2"
673..675
/gene="dif-2"
/citation={3}
/replace="ag"
687..689
/gene="dif-2"
/citation={3}
/replace="tg"
783..785
/gene="dif-2"
/citation={3}
/replace="tg"
1020..1021
/gene="dif-2"
/citation={3}
/replace="att"
1120
/gene="dif-2"
/citation={3}
/replace="g"
1133
/gene="dif-2"
/citation={3}
/replace="t"

BASE COUNT      259 a      367 c      342 g      262 t
ORIGIN

Query Match      100.0%; Score 258; DB 53; Length 1230;
Best Local Similarity 100.0%; Pred. No. 6.8e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGTGCAGTGCAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
DB 222 GTCGGCGCCAGTGCAGTGCAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 281
QY 61 CTCACCATCGTCTTCTGCGAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120
DB 282 CTCACCATCGTCTTCTGCGAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 341
QY 121 CCAGAGGAGCCCTTAACCCCGCATCCCTGCGGCCACCCCTGTCGCCCGCCCTGCCT 180
DB 342 CCAGAGGAGCCCTTAACCCCGCATCCCTGCGGCCACCCCTGTCGCCCGCCCTGCCT 401
QY 181 CCGTTTAACTGACTTCGGAGCCCTCGGACTACGGCTCTGGACCTCAGACACTTCTCTCCAG 240
DB 402 CCGTTTAACTGACTTCGGAGCCCTCGGACTACGGCTCTGGACCTCAGACACTTCTCTCCAG 461

```

```

QY 241 CAACACCGCGCCCTTC 258
DB 462 CAACACCGCGCCCTTC 479

RESULT 3
AF039067 1309 bp mRNA PRI 21-AUG-1998
LOCUS Homo sapiens anti-death protein (IEX-1L) mRNA, complete cds.
DEFINITION AF039067
VERSION AF039067.1 GI:3449375
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Wu.M.X., Ao.Z., Prasad,K.V.S., Wu.R. and Schlossman,S.F.
TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell
survival
JOURNAL Science 281 (5379), 998-1001 (1998)
MEDLINE 98369175
REFERENCE 2 (bases 1 to 1309)
AUTHORS Ao.Z. and Wu.M.X.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Tumor Immunology, Dana Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
Location/Qualifiers
1..1309
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1309
/gene="IEX-1L"
1..582
/gene="IEX-1L"
/feature="type II membrane protein"
/codon_start=1
/product="anti-death protein"
/protein_id="AAC32558.1"
/db_xref="GI:3449376"
/translation="MCHSRSCPTMTILQAPTAPSTIPGPRRSGPPTFTDPLPEP
AAPAGRPSASGRHKRRSRVLYPRVRRQLPVEENPAKRLLELLLTIVFCQILMAE
EGVAPALPPEDAPNAASLAPTVPVSPVLEPNTSEPSYALDLSFLQOHPAF"
BASE COUNT      262 a      393 c      369 g      285 t
ORIGIN

```

```

Query Match      100.0%; Score 258; DB 10; Length 1309;
Best Local Similarity 100.0%; Pred. No. 6.7e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGTGCAGTGCAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
DB 322 GTCGGCGCCAGTGCAGTGCAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 381
QY 61 CTCACCATCGTCTTCTGCGAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120
DB 382 CTCACCATCGTCTTCTGCGAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 441
QY 121 CCAGAGGAGCCCTTAACCGCCATCTTGGCGCCACCCCTGTCGCCCGCCCTGCCT 180
DB 442 CCAGAGGAGCCCTTAACCGCCATCTTGGCGCCACCCCTGTCGCCCGCCCTGCCT 501
QY 181 CCGTTTAACTGACTTCGGAGCCCTTCGGACTACGGCTCTGGACCTCAGACACTTCTCTCCAG 240
DB 502 CCGTTTAACTGACTTCGGAGCCCTTCGGACTACGGCTCTGGACCTCAGACACTTCTCTCCAG 561
QY 241 CAACACCGCGCCCTTC 258
DB 562 CAACACCGCGCCCTTC 579

```

Matches 258: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCCGAGCTCCAGTCGAGAACCCAGACCCAAAGGCTTCTTTCTGCTG 60
 DB 914 GTCCGGCCGAGCTCCAGTCGAGAACCCAGACCCAAAGGCTTCTTTCTGCTG 973
 QY 61 CTCACCATCTCTCTCCAGATCCTGATGGCTGAAGAGGTGTGCGGGCCGCTGCCT 120
 DB 974 CTCACCATCTCTCTCCAGATCCTGATGGCTGAAGAGGTGTGCGGGCCGCTGCCT 1033
 QY 121 CCAGAGAGCCCTATAGCGGGATCCTGCGGCCACCCCTGTGTGCCCGTCTCGAG 180
 DB 1034 CCAGAGAGCCCTATAGCGGGATCCTGCGGCCACCCCTGTGTGCCCGTCTCGAG 1093
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGACCTCAGCACTTCTCCAG 240
 DB 1034 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGACCTCAGCACTTCTCCAG 1153
 QY 241 CAACACCGCGCCCTTC 258
 DB 1154 CAACACCGCGCCCTTC 1171

RESULT 6

AC006165 44118 bp DNA PRI 08-DEC-1998
 LOCUS Homo sapiens clone UMG:Y54C125 from 6p21, complete sequence.
 DEFINITION AC006165
 ACCESSION AC006165
 VERSION AC006165.1 GI:3980464
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 44118)
 AUTHORS Janet, M., Guillaudoux, T., Vu, Q., Kutayavin, T., Harter, H. and Geraghty, D.E.
 TITLE Large scale sequence analysis of the human MHC class I region
 JOURNAL Unpublished (1998)
 REMARK Fred Hutchinson Cancer Research Center
 The Clinical Research Division
 1100 Fairview Ave. N., P.O. Box 19024
 Seattle, WA 98109-1024
 REFERENCE 2 (bases 1 to 44118)
 AUTHORS Geraghty, D.E. and Olson, M.V.
 TITLE Direct Submission
 JOURNAL Submitted (08-DEC-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 REMARK University of Washington Human Genome Center
 Box 352145, Seattle, WA 98195
 Contact: Daniel E. Geraghty (geraghty@fhcrc.org)
 Overlapping Sequences:
 5': UMG:Y54C283
 3': UMG:Y54C222 (Genbank Accession: AC006049)

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

 Double stranded (DS) coverage: 58.8%
 DS or two chemistry coverage: 99.5%
 Single stranded regions: 2

 Sequence Validation:
 This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below.

Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

Map	Seq	HindIII	Map	Seq	NsiI
6714..50	6452..00	1823..23	1794..00	5581..12	5401..00
1089..04	1085..00	1246..04	1228..00	10570..24	10123..00
1536..15	1528..00	5570..33	5372..00	10069..29	9626..00
3623..08	3537..00	1670..92	1630..00		
3784..08	3621..00	1270..76	1280..00		
6834..00	6689..00	4235..79	4109..00		
11234..36	10874..00	1171..00	1151..00		
762..43	762..00	3226..11	3096..00		
2727..38	2653..00	10501..64	10080..00		
1691..42	1684..00	4208..44	4052..00		
1295..60	1277..00				
746..87	734..00				

FEATURES		Location/Qualifiers			
Source		1..44118			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/chromosome="6"			
		/map="6p21"			
		/clone="CGM1.D100E8"			
		/sub_clone="UMG:Y54C125"			
		/cell_line="CGM1"			
		/clone_lib="Wash U YAC library"			
		218..405			
repeat_region		/rpt_family="MER3"			
		400..999			
repeat_region		/rpt_family="Alu"			
		1797..2378			
repeat_region		/rpt_family="Alu"			
		2649..2956			
repeat_region		/rpt_family="Alu"			
		complement(3372..3894)			
repeat_region		/rpt_family="Alu"			
		3941..4070			
repeat_region		/rpt_family="Alu"			
		complement(4449..4724)			
repeat_region		/rpt_family="Alu"			
		complement(5117..5389)			
repeat_region		/rpt_family="Alu"			
		7668..7913			
repeat_region		/rpt_family="Alu"			
		complement(7924..8226)			
repeat_region		/rpt_family="Alu"			
		8250..8551			
repeat_region		/rpt_family="Alu"			
		complement(8893..8973)			
repeat_region		/rpt_family="MIR"			
		complement(11829..12096)			
repeat_region		/rpt_family="Alu"			
		12416..12740			
repeat_region		/rpt_family="Alu"			
		complement(13181..13247)			
repeat_region		/rpt_family="Alu"			
		complement(13498..13647)			

COMMENT

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 0
Center clone name: RPCI-23_128D3

Summary Statistics
Consensus quality: 139878 bases at least Q40
Consensus quality: 165603 bases at least Q30
Consensus quality: 177139 bases at least Q20
Estimated insert size: 183300; agarose-fp estimation
Estimated insert size: 200576; sum-of-contigs estimation
Quality coverage: 3.48 in Q20 bases; agarose-fp estimation
Quality coverage: 3.18 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1175: contig of 1175 bp in length
1176: gap of unknown length
1276: contig of 1288 bp in length
2563: gap of unknown length
2564: gap of unknown length
3908: contig of 1243 bp in length
4008: gap of unknown length
5224: contig of 1218 bp in length
5225: gap of unknown length
5325: contig of 1312 bp in length
6636: gap of unknown length
6737: contig of 2460 bp in length
9198: gap of unknown length
9296: gap of unknown length
10723: contig of 1427 bp in length
10724: gap of unknown length
10824: contig of 2029 bp in length
12853: gap of unknown length
12953: contig of 3073 bp in length
16026: gap of unknown length
16126: gap of unknown length
16128: contig of 2971 bp in length
19098: gap of unknown length
19196: gap of unknown length
21750: contig of 2554 bp in length
21751: gap of unknown length
21850: gap of unknown length
21851: contig of 4729 bp in length
26579: gap of unknown length
26580: gap of unknown length
32478: contig of 5793 bp in length
32479: gap of unknown length
32573: gap of unknown length
36994: contig of 4422 bp in length
37094: gap of unknown length
44071: contig of 6977 bp in length
44171: gap of unknown length
50721: contig of 6556 bp in length
50827: gap of unknown length
57204: contig of 6377 bp in length
57304: gap of unknown length
65076: contig of 7772 bp in length
65176: gap of unknown length
71824: contig of 6648 bp in length
71924: gap of unknown length
78035: contig of 6111 bp in length
78135: gap of unknown length
87838: contig of 9703 bp in length
87839: gap of unknown length
87939: contig of 6986 bp in length
95024: gap of unknown length
95025: contig of 6164 bp in length
101189: gap of unknown length
101289: contig of 7637 bp in length
108925: gap of unknown length
109025: gap of unknown length

* 109026 119738: contig of 10713 bp in length
* 119739 119838: gap of unknown length
* 119839 129876: contig of 10038 bp in length
* 129877 129976: gap of unknown length
* 129977 143243: contig of 13267 bp in length
* 143244 143343: gap of unknown length
* 143344 161164: contig of 17821 bp in length
* 161165 161264: gap of unknown length
* 161265 180316: contig of 19032 bp in length
* 180317 180416: gap of unknown length
* 180417 203476: contig of 23060 bp in length.
FEATURES
Location/Qualifiers
1..203476
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="Rp23-128D3"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 51748 a 48714 c 48138 g 51580 t 3296 others
ORIGIN

Query Match 58.8%; Score 151.2; DB 70; Length 203476;
Best Local Similarity 75.9%; Pred. No. 4.7e-24;
Matches 205; Conservative 0; Mismatches 53; Indels 12; Gaps 1;
QY 1 GTCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCAAGGCTCTCTTTCTGCTG 60
|||||
Db 92875 GTCGGCGCCAGCTACCACCGAGGAACCAACATTCGCAAGAGGGTCTCTTTCTGCTG 92934
QY 61 CTCACCATCGTCTTCTCCAGATCCTGATGGCTGAAGAGGCTGTGCGGGGCGCCCTGCCT 120
|||||
Db 92935 TTGCCATCATCTTCTCCAGATTTTCATGGCTGACAGAGGCTGTGTCGACGCCCTGGCT 92994
QY 121 CCAGAGACGCCCTACGCGGCATCCC-----TGGCGCCACCCCTGTGTCC 168
|||||
Db 92995 CCGAGGATGCTACCGCGCGTGCACACCTGAGCCCATTTCTGCGCCCATTAATCTCGCCC 93054
QY 169 CCCTCTCGAGCCCTTTAATCTGACTTCGAGGCCCTCGGACTACGCTCTGGACCTCAGC 228
|||||
Db 93055 CCGTCTCGAGCCCTTGACCTGACCTCGAGTCTCGGACTCTCGGACTATGCGTGGATCTTAAA 93114
QY 229 ACTTCTCCAGCAACACCGCGCGCTTC 258
|||||
Db 93115 GCTTTCTCCAGCAACATCGCGCGCTTC 93144

RESULT 12

MMGLY96 1938 bp mRNA ROD 11-MAR-1993
LOCUS M.musculus gly96 mRNA.
DEFINITION X67644
ACCESSION X67644
VERSION X67644.1 GI:287803
KEYWORDS gly96 gene; glycosylated protein.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1938)
AUTHORS Lau, L.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1992) L. Lau, University of Illinois College of
Med., Dept. of Genetics, M/C 669, 808 South Wood Street, Chicago,
IL 60612, USA
REFERENCE 2 (bases 1 to 1938)
AUTHORS Charles, C.H.; Yoon, J.K.; Simske, J.S. and Lau, L.F.
TITLE Genomic structure; cDNA sequence, and expression of gly96, a growth
factor-induced immediate-early gene encoding a short-lived
glycosylated protein
JOURNAL Oncogene 8 (3), 797-801 (1993)
MEDLINE 93173526
FEATURES Location/Qualifiers
1..1938
/organism="Mus musculus"

